Final Report Briefing
Analytic Method for Cost and Schedule Risk Analysis
March 4, 2013

presented by:

Raymond Covert

rpcovert@covarus.com

Acknowledgements

- The author would like to acknowledge the following individuals:
  - NASA Office of Program Analysis and Evaluation (PA&E), Cost Analysis Division (CAD) for this research. Specific thanks to: Mr. Charles Hunt, Dr. William Jarvis and Mr. Ronald Larson
  - My late mentor, Dr. Stephen Book, whose work laid the foundations of this report
  - Dr. Paul Garvey (MITRE) and Mr. Timothy Anderson (iParametrics) for their assistance with many of the difficult subjects approached in the report
  - Galorath, Inc.: Mr. Dan Galorath for his gracious assistance in making this work possible; Mr. Robert Hunt for his diligence as the project manager; Mr. Brian Glauser for his promotion of the effort; Ms. Wendy Lee for her help deciphering the elusive properties of the beta distribution; and Ms. Karen McRitchie for her support in motivating the technical application of the analytic method into SEER-H

Agenda

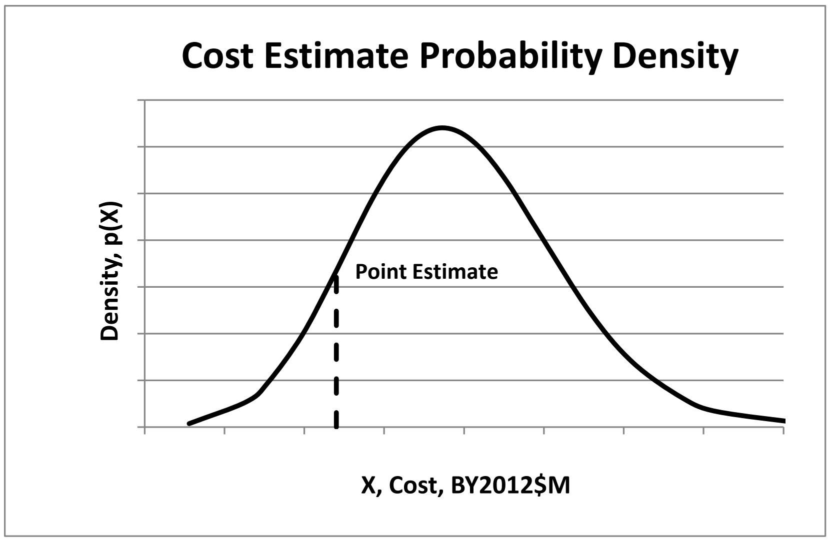
- Introduction
- Mathematical Problems in Cost and Schedule Estimates
- Probability Tools
- Products of Random Variables
    - Expectation Methods
    - Propagation of Errors
- Functional Correlation
- Discrete Risks
- Max and Min of Random Variables
- Parametric Estimate Example Problem
- Resource-Loaded Schedule Example Problem
- Summary
- Future Research

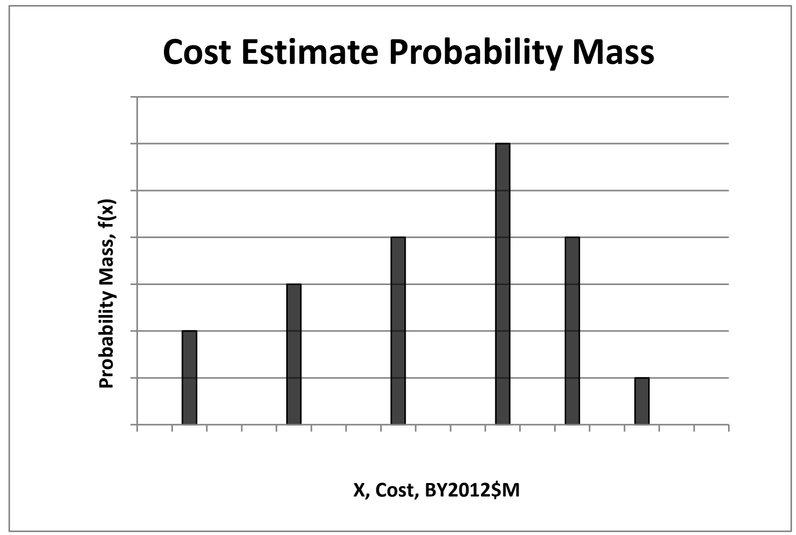
Introduction

- Cost and schedule estimates are probabilistic in nature
- We do not know the <u>exact</u> cost or schedule duration of a project until it is complete and we have collected "actual" cost and schedule durations, so
  - Discrete numbers are not a good representation of expected cost or schedule duration – they are bound to be incorrect
  - Until the project is complete, we must rely on estimates
- Estimates imply uncertainty, and the mathematics of uncertainty is probability,
- …so estimates <u>must</u> be expressed probabilistically (i.e., as probability distributions)
  - The difficult part is finding a solution to the mathematical problem of the estimate!

Probability Distributions (1)

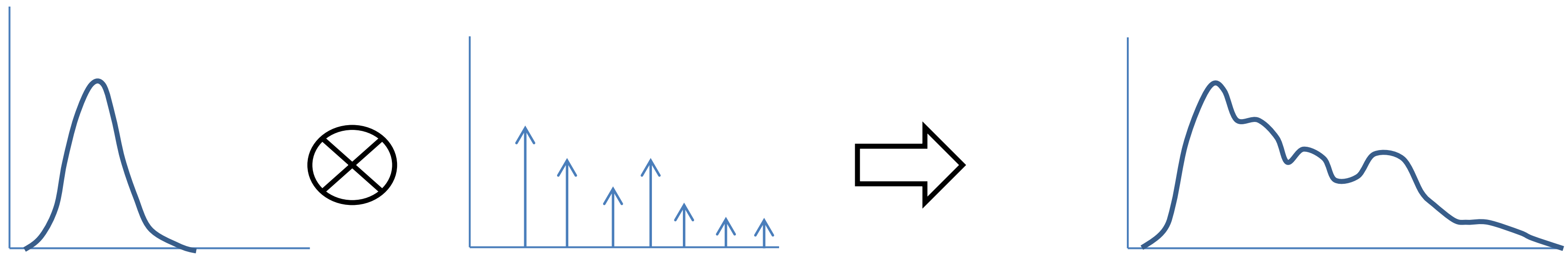
- A probability distribution completely defines a random variable
  - Continuous Probability Distribution (density)
  - Discrete Probability Distribution (mass)
  - Mixed (or mixture) Probability Distribution (mass and density)





Probability Distributions (2)

- Mixed (or "mixture") distribution formed by *convolution* of continuous and discrete distributions



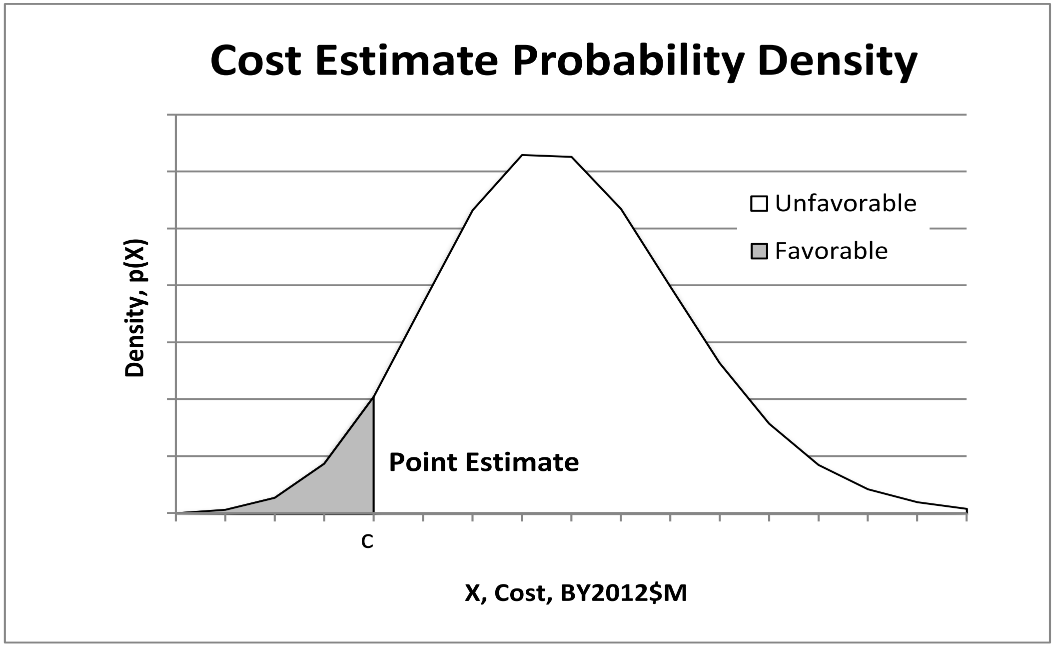
CONTINUOUS DISTRIBUTION

DISCRETE DISTRIBUTION

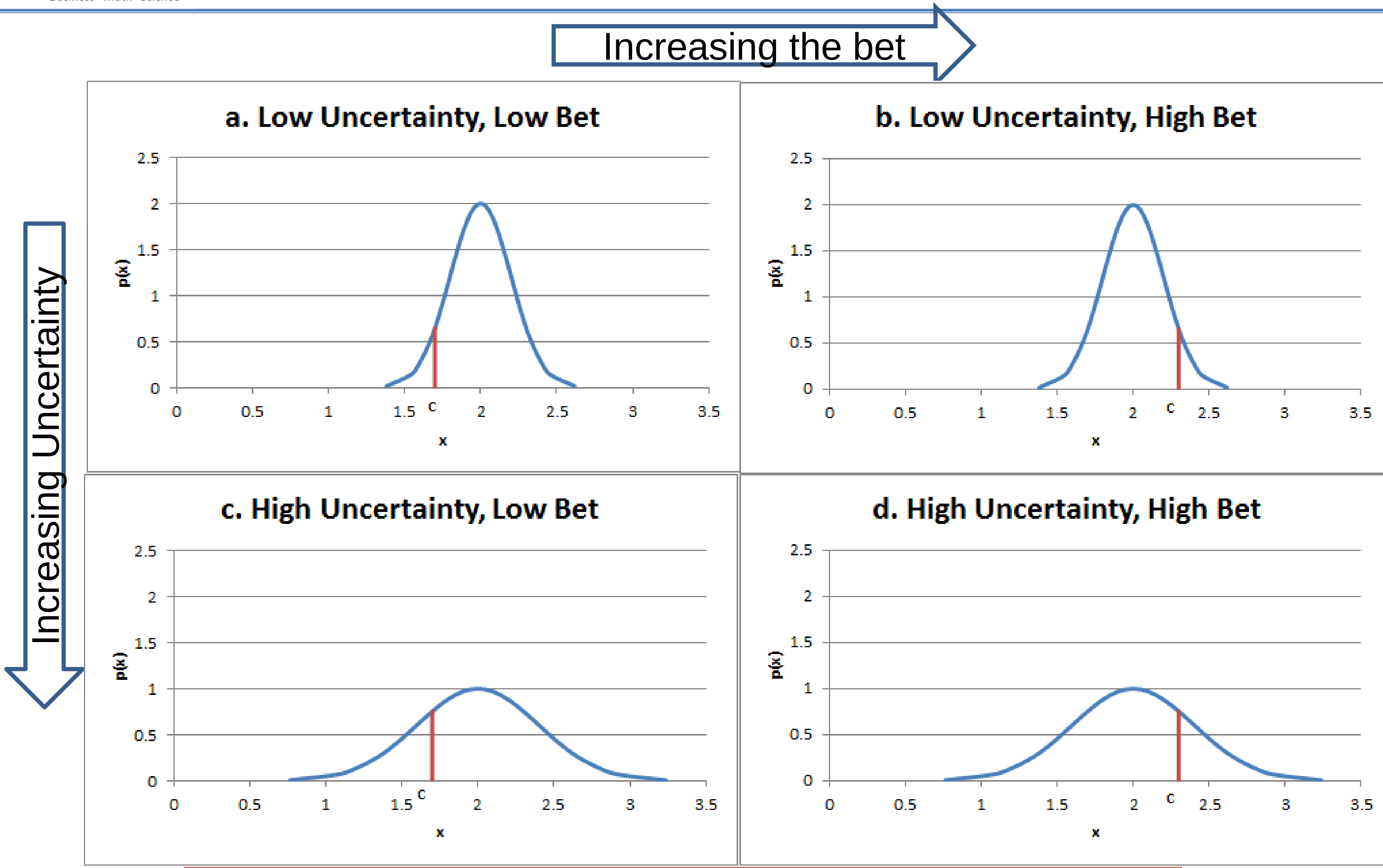
MIXED DISTRIBUTION

Risk, Uncertainty and the Bet (1)

- Uncertainty is defined by the estimate's probability distribution
- The "bet", shown as "c", is a discrete point estimate
- Risk and opportunity are defined by the areas of the distribution to the left (opportunity) and right (risk) of the bet



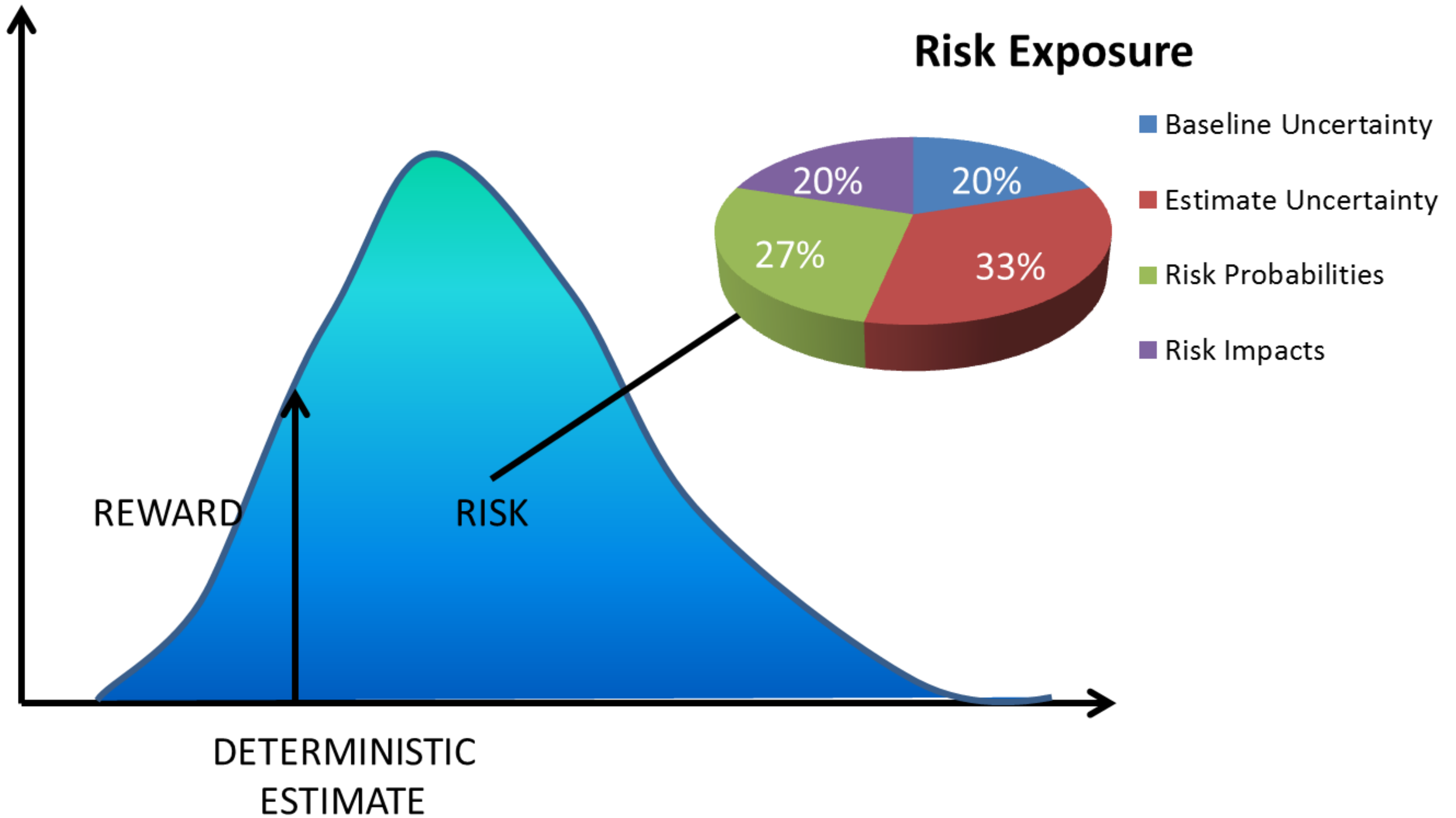
Risk, Uncertainty and the Bet (2)



Changing the bet or distribution will change the risk!

Risk Components

- If we know the bet and components of the probability distribution of the estimate, we can segregate risk drivers



Joint Probability Distributions

- If we have two random variables X and Y, we can define the probabilities

$$P\{X \le x\} = F_X(x) = \int_{-\infty}^{x} F_X(z)\,dz$$

$$P\{Y \le y\} = F_Y(y) = \int_{-\infty}^{y} F_Y(z)\,dz$$

- The joint probabilities of $P\{X \le x, Y \le y\}$ can be expressed as the joint distribution function

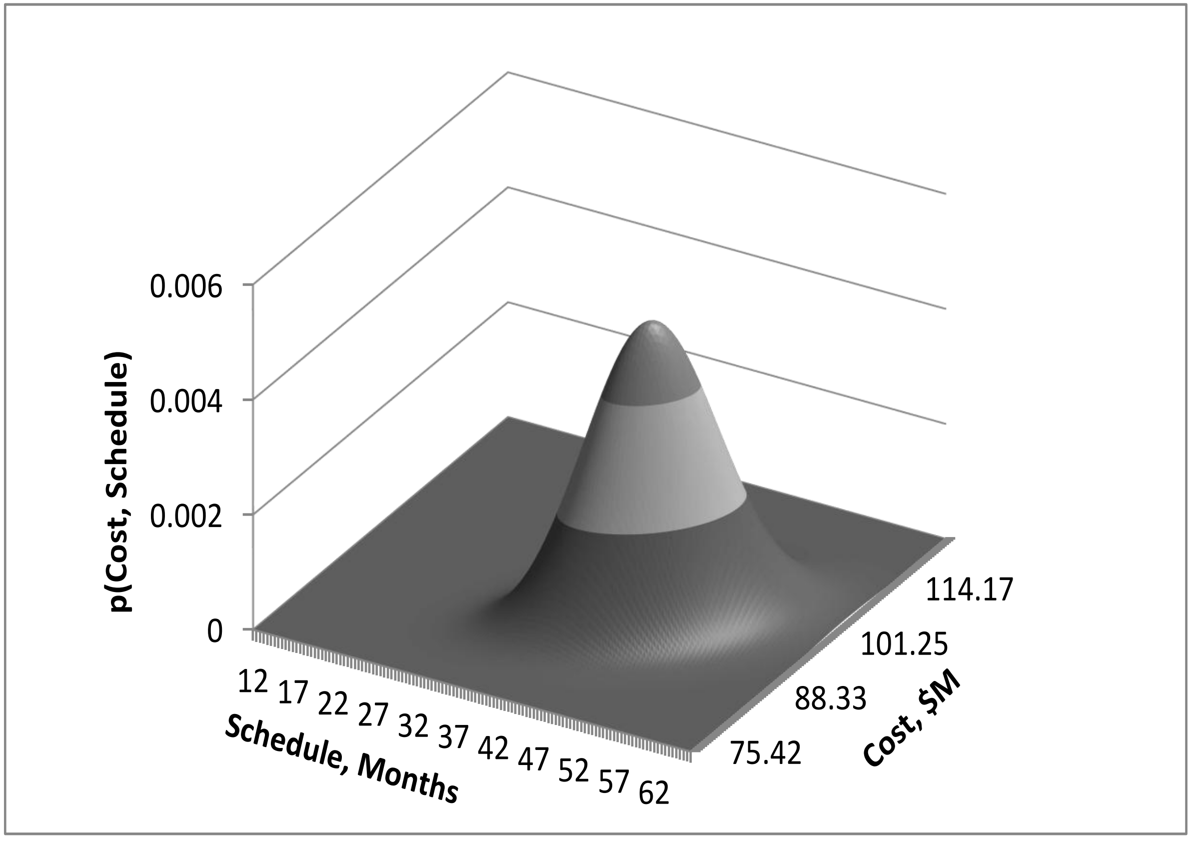
$$P\{X \le x, Y \le y\} = F_{XY}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{XY}(z,w)\,dz\,dw$$

- And the joint probability density function (PDF) is defined as

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}$$

Joint Cost and Schedule Probability Distribution

- This is a joint PDF of cost and schedule generated with Excel



Statistics of a Random Variable

- Moments
  - Provide useful information about the characteristics of a random variable, X, such as the measures of central tendency, dispersion and shape
    - Always define which of the three types of moments being used: raw moments, central moments or standardized moments
  - Raw Moments
    - About the origin (i.e., zero)
  - Central Moments
    - About the mean, μ, or first raw moment
  - Standardized Moments
    - Normalized by dividing by the k^{th} power of the standard deviation, σ
- Order Statistics
  - Maximum, minimum, first, last, etc.

Raw Moments

- The k^{th} moments about the origin are called "raw moments" of a PDF, f_X, and are defined as:

$$\mu'_k = \begin{cases} \sum_X x^k f(x) & ; if\ X\ is\ discrete \\ \int_{-\infty}^{\infty} x^k f(x) dx & ; if\ X\ is\ continuous \end{cases}$$

- The mean, μ'_1 , is the first raw moment of X about the origin, and it is a measurement of the central tendency of the data
- We are more familiar with the mean being represented as, μ, so we will use this notation for the mean hereafter

Central Moments (1)

- Central moments of a distribution are the raw moments about the mean, μ
- The first central moment is, by definition zero, but the second central moment is the variance, σ^2, which is a measure of dispersion about μ
- Definition of the k^{th} central moments of discrete and continuous random variables (RVs)

$$\sigma^2 = \begin{cases} \displaystyle\sum_X (x-\mu)^k f(x) & ; if\ X\ is\ discrete \\ \displaystyle\int_{-\infty}^{\infty} (x-\mu)^k f(x)dx & ; if\ X\ is\ continuous \end{cases}$$

Central Moments (2)

- The first five *central moments* expressed in terms of the raw moments are

$$\mu_1 = 0$$

$$\mu_2 = -{\mu'_1}^2 + \mu'_2 = \mu'_2 - {\mu'_1}^2 \longleftarrow \text{Variance, } \sigma^2$$

$$\mu_3 = 2{\mu'_1}^3 - 3\mu'_1\mu'_2 + \mu'_3$$

$$\mu_4 = -3{\mu'_1}^4 + 6{\mu'_1}^2\mu'_2 - 4\mu'_1\mu'_3 + \mu'_4$$

$$\mu_5 = 4{\mu'_1}^5 - 10{\mu'_1}^3\mu'_2 + 10{\mu'_1}^2\mu'_3 - 5\mu'_1\mu'_4 + \mu'_5$$

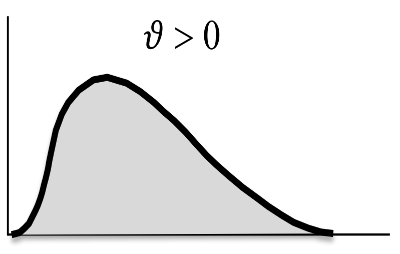
Standardized Moments (1)

- Standardized moments are the kth central moments, μ_k, normalized by the kth powers of the standard deviation σ^k (i.e., $\frac{\mu_k}{\sigma^k}$)
- Skewness, ϑ, is the measure of asymmetry of X and is defined as the third standardized moment

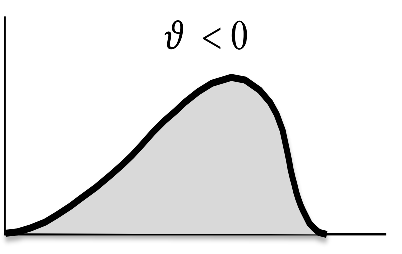
$$skew(X) = \vartheta = \frac{\mu_3}{\sigma^3}$$

- A distribution is a) symmetric if $\vartheta = 0$, b) left (i.e. negatively) skewed if $\vartheta < 0$, and c) right (i.e., positively) skewed if $\vartheta > 0$

Right Skewed



Left Skewed



Standardized Moments (2)

- Kurtosis is the fourth standardized moment
  - Most textbooks define kurtosis of *symmetric, unimodal* distributions as a measure of peakedness of a distribution X
  - This is a correct definition, however a more descriptive definition of kurtosis exists - the measure of the dispersion around the two "shoulders" of a distribution located at $\mu \pm \sigma$
  - The classical attribution of peakedness of a distribution vice its "fat-tailedness" is not a good representation of the meaning of kurtosis

$$kurt(X) = \frac{\mu_4}{\sigma^4}$$

- A more commonly used metric is the "excess kurtosis", which is $kurt(X) - 3$

$$\kappa = kurt(X) - 3 = \frac{\mu_4}{\sigma^4} - 3$$

Expectation

- The expectation, $E[\cdot]$, of a random variable is a powerful expression
- The expected value, or μ, of a random variable is perhaps the most important single parameter in applied probability
- It is written as $E[X] = \mu_X$, and is the integral $E[X] = \int_{-\infty}^{\infty} x f_X(x)dx$, where $f_X(x)$ is the PDF of X
  - Same as the raw moment (the probability center of mass)
- Another important parameter is σ^2, defined by the expectation of the squared difference of the PDF and its mean
  - This quantity represents the moment of inertia of the probability masses

$$\mathrm{Var}(X) = \sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (X-\mu)^2 f_X(x)dx$$

The Importance of E[·]

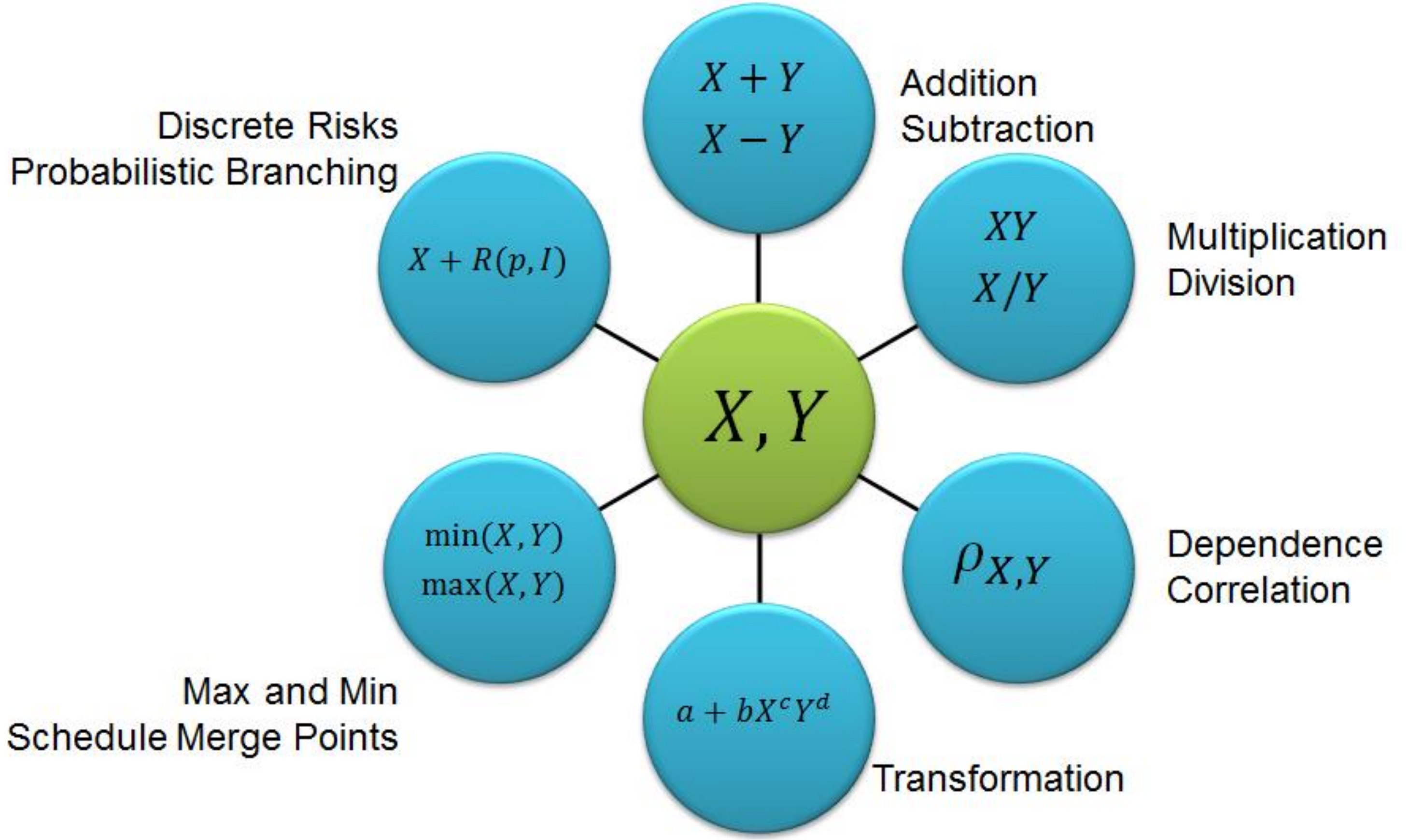
- What is most important about $E[\cdot]$ is its ability to determine the raw moments and central moments of a random variable, and thus the measures of central tendency, dispersion and shape (i.e., $\mu, \sigma^2, \vartheta, \kappa$)
- If we can calculate the expectation, we can generate the moments
- If we know the moments we can figure out the approximate shape of the distribution

Agenda

- Introduction
- Mathematical Problems in Cost and Schedule Estimates
- Probability Tools
- Products of Random Variables
  - Expectation Methods
  - Propagation of Errors
- Functional Correlation
- Discrete Risks
- Max and Min of Random Variables
- Parametric Estimate Example Problem
- Resource-Loaded Schedule Example Problem
- Summary
- Future Research

Mathematics of Estimates

- Cost and schedule estimates rely on these operations of random variables (X and Y)
  - Which provide information used in uncertainty and risk analysis



Agenda

- Introduction
- Mathematical Problems in Cost and Schedule Estimates
- Probability Tools
- Products of Random Variables
  - Expectation Methods
  - Propagation of Errors
- Functional Correlation
- Discrete Risks
- Max and Min of Random Variables
- Parametric Estimate Example Problem
- Resource-Loaded Schedule Example Problem
- Summary
- Future Research

Probability Tools

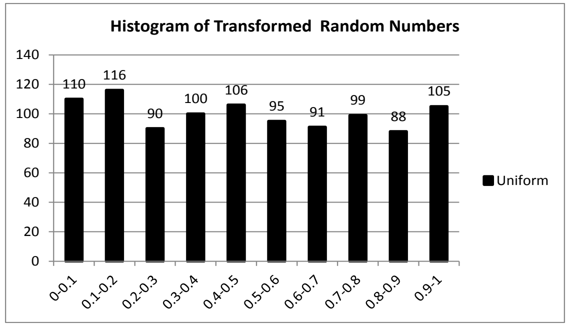
- When we perform a cost or schedule risk analysis, we need to know the uncertainty of the individual estimates, how they are correlated, and how to combine them
  - In a work breakdown structure (WBS)
  - In a schedule network
- We can employ statistical modeling techniques such as statistical simulation or statistical analysis to find these uncertainties and their properties
- Although the goal is the same, these techniques differ, which we will discuss in more detail

Statistical Simulation

- The statistical simulation process follows these steps:
  1. Define numerical experiment (spreadsheet, schedule network, etc.)
  2. Define PDFs for each random variable
  3. Define correlation coefficients between random variables
  4. Determine the number of experimental trials
  5. For each trial:
     1. Draw correlated random variable(s) from defined PDF(s)
        1. Sample uniform distributions, $U(1,0)$
        2. Transform each $U(1,0)$ to the desired PDF based on an inverse transformation of the cumulative distribution (CDF), CDF^{-1}.
        3. Correlate the set of PDFs
     2. Compute the experimental result(s)
     3. Save the experimental result(s)
  6. At the end of the simulation, determine the statistics from the experimental results

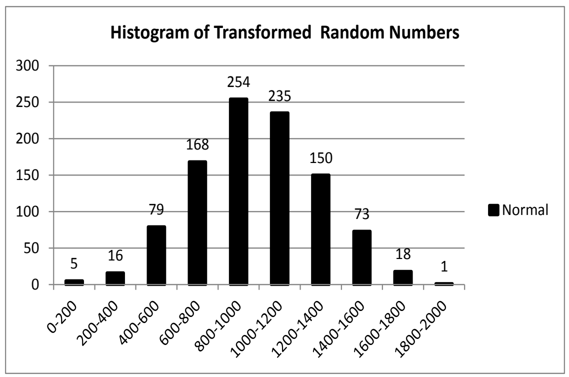
Example Sampled Distributions

- Uniform



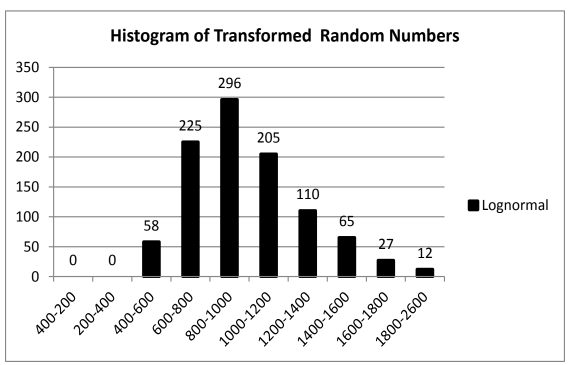
Moment	Simulated	Exact
μ	0.488	0.500
σ	0.292	0.083
ϑ	0.053	0.000
κ	-1.222	-1.200

- Normal



Moment	Simulated	Exact
μ	987.7155	1000
σ	303.4236	300
ϑ	0.001349	0
κ	-0.12993	0

- Lognormal



Moment	Simulated	Exact
μ	988.989	1000
σ	299.102	300
ϑ	0.855934	0.927
κ	1.094075	1.566

Benefits and Drawbacks of Statistical Simulation

- Among its benefits are…
  - its ability to provide the statistics of a simulated PDF formed by complex mathematical modeling of random variables
  - its relative ease of use
- Most of the time, statistical simulation obtains very close results to [and is easier to use than] statistical analysis
- However, statistical simulation does have its drawbacks
  - inability to sample uniformly
  - (in)ability to correlate two distributions <u>exactly</u> using Pearson product-moment correlation coefficients
  - inability to sample INDEPENDENT distributions (only uncorrelated)
  - inability to correlating large numbers of random variables, and
  - inability to provide reasonable results when the number of simulation trials is too small to account for single or combinations of low-probability events (the tails)

Statistical Analysis

- Unlike simulation, statistical analysis relies on the exact calculation of moments of the PDF
- We use moments as the basis of the analytical technique proposed in this report
- Method of Moments
  - Method of Moments (MOM) is a relatively easy-to-use, analytical technique used to calculate the moments of probability distributions
  - Relies on statistical calculations of moments to derive the statistics of probability distributions such as WBS element cost estimates or schedule durations
  - With the widespread use of statistical simulation tools by cost and schedule analysts, MOM has become a forgotten “art”
  - One of the surviving MOM techniques is the Formal Risk Assessment of System Cost Estimates (FRISK) method

FRISK

- The FRISK method is a MOM approach used to calculate the μ and σ^2 of the PDF of total cost formed by the statistical summation of PDFs of subordinate cost elements
- It relies on the calculation of moments of WBS elements defined as triangular distributions and statistically sums them to find the parameters of an assumed lognormal PDF
- A very useful, quick-look tool that can be programmed in just a few minutes
- The major drawback is the assumption of distribution shapes
  - We can make provisions for this by examining higher moments such as skewness and kurtosis

FRISK Process

1. Define numerical experiment; in this case, the summation of a WBS
2. Define triangular PDFs, $T(L_i, M_i, H_i)$ for each cost, X_i , or random variable to be statistically summed, by specifying the low (L_i), most likely (M_i) and high (H_i) values
3. Calculate the μ_i and $\sigma_i{}^2$ for each $T(L_i, M_i, H_i)$
4. Sum the n means to calculate the mean of the sum of the PDFs
5. Define correlation coefficients, $\rho_{i,j}$, for each pair of PDFs
6. Calculate the total variance of the sum of the PDFs
7. Assume the PDF of the total cost is a lognormal distribution, $L(P, Q)$
8. Calculate the lognormal parameters P and Q
9. Determine the percentile statistics $L(P, Q)_Z$ using the inverse CDF tables or the *LOGINV* function in Excel

FRISK Example (1)

- Here is a set of example inputs

WBS Element, i	L_i	M_i	H_i
Antenna	191	380	1151
Electronics	96	192	582
Platform	33	76	143
Facilities	9	18	27
Power Distribution	77	154	465
Computers	30	58	86
Environmental Control	11	22	66
Communications	58	120	182
Software	120	230	691
TOTAL	625	1250	3393

Point Estimate

- This is the matrix of correlation coefficients between each WBS element

$$\begin{bmatrix} 1 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 1 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.2 & 1 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.2 & 0.2 & 1 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 1 & 0.2 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 1 \end{bmatrix}$$

FRISK Example (2)

- These are the resulting moment calculations

$$\sigma_{f(x)_1} = \sqrt{\sigma_{f(x)_1}{}^2} =$$

$$\sqrt{\frac{[191^2+380^2+1151^2-(191)(380)-(191)(1151)-(380)(1151)]}{18}} = \$207.62K$$

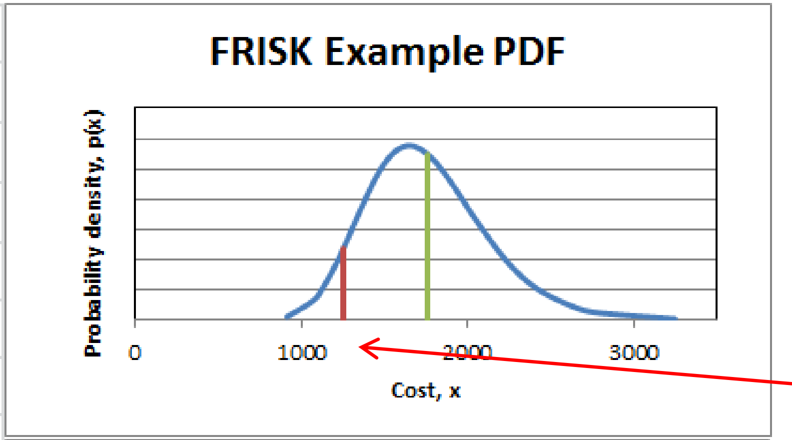
$$\mu_1 = (L_1 + M_1 + H_1)/3 = \frac{191+380+1151}{3} = \$574K$$

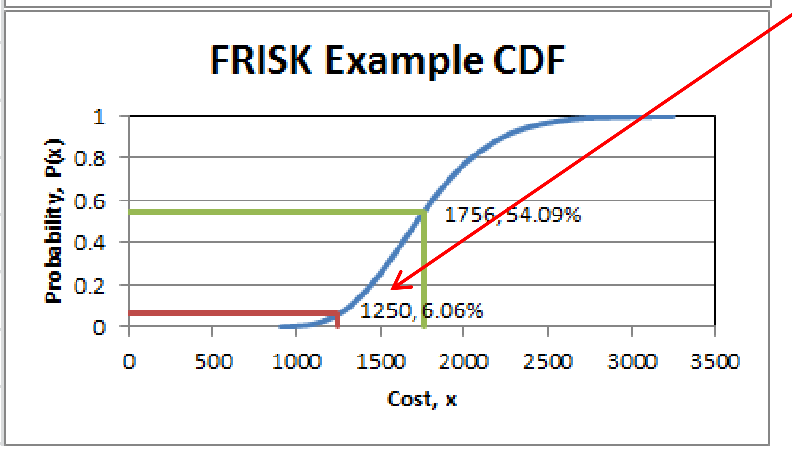
WBS Element, i	Estimate, $f(x)_i$	$\mu_{f(x)_i}$	$\sigma_{f(x)_i}$
Antenna	T(191,380,1151)	574	207.62
Electronics	T(96,192,582)	290	105.08
Platform	T(33,76,143)	84	22.63
Facilities	T(9,18,27)	18	3.67
Power Distribution	T(77,154,465)	232	83.86
Computers	T(30,58,86)	58	11.43
Environmental Control	T(11,22,66)	33	11.88
Communications	T(58,120,182)	120	25.31
Software	T(120,230,691)	347	123.68
TOTAL (Not necessarily the sum)		1756	364.93

FRISK Example (3)

- From this information, we can plot the PDF and cumulative density function (CDF) and report relevant statistics

Percentile	Value
10%	1320.981
20%	1446.051
30%	1543.52
40%	1631.993
50%	1719.265
60%	1811.205
70%	1915.021
80%	2044.1
90%	2237.635
mean	1756.000
median	1719.265
mode	1648.086
std. dev.	364.9331





Agenda

- Introduction
- Mathematical Problems in Cost and Schedule Estimates
- Probability Tools
- Products of Random Variables
  - Expectation Methods
  - Propagation of Errors
- Functional Correlation
- Discrete Risks
- Max and Min of Random Variables
- Parametric Estimate Example Problem
- Resource-Loaded Schedule Example Problem
- Summary
- Future Research

Product of RVs

- One of the most difficult parameters to estimate is the variance of the product of two dependent RVs ($Z = XY;\ \rho_{XY} \neq 0$)
  - The mean of Z is found by this (fairly straightforward) equation
    $$\mu_Z = \mu_X \mu_Y + \rho_{XY} \sigma_X \sigma_Y$$
  - The variance is not
- We had to research the variance of the product of normal and lognormal RVs which provided inconsistent formulae
  - We then derived the equation from scratch to see where the inconsistencies arose
  - Comparison with statistical simulation tool showed inconsistent results (from the simulation tool) so we had to discover why…

Variance of a Product of Two RVs

- From Goldberger

$$Var[XY] = \rho_{X^2,Y^2}\sigma_{X^2}\sigma_{Y^2} + \sigma_X{}^2\mu_Y{}^2 + \sigma_Y{}^2\mu_X{}^2 + \sigma_X{}^2\sigma_Y{}^2 - [2\rho_{X,Y}\sigma_X\sigma_Y\mu_X\mu_Y + (\rho_{X,Y}\sigma_X\sigma_Y)^2]$$

- If they are truly independent, then

$$Var[XY] = \sigma_X{}^2\mu_Y{}^2 + \sigma_Y{}^2\mu_X{}^2 + \sigma_X{}^2\sigma_Y{}^2 \text{, and } \sigma_{XY} = \sqrt{\sigma_X{}^2\mu_Y{}^2 + \sigma_Y{}^2\mu_X{}^2 + \sigma_X{}^2\sigma_Y{}^2}$$

- If they are not, which is often the case, then we have problems
- The first problem is finding the value of the term $\rho_{X^2,Y^2}\sigma_{X^2}\sigma_{Y^2}$
  - What is the covariance of the squares of two RVs?
  - $\rho_{X^2,Y^2}\sigma_{X^2}\sigma_{Y^2} = Cov(X^2, Y^2) = E((X^2Y^2)^2) - E(X^2)\,E(Y^2)$
  - When the two RVs are normally-distributed then the problem is easier to solve

Product of Two Normal PDFs

- If X and Y are bivariate normally distributed then the third moments vanish and the term $E\left((X^2Y^2)^2\right) - E(X^2)\ E(Y^2)$ reduces to ${\sigma_X}^2\sigma_{Y_2^2} + 2Cov(x,y)^2$
- This means $Var[XY] = {\sigma_X}^2{\mu_Y}^2 + {\sigma_Y}^2{\mu_X}^2 + {\sigma_X}^2{\sigma_Y}^2 + 2\rho_{X,Y}\sigma_X\sigma_Y\mu_X\mu_Y + \left(\rho_{X,Y}\sigma_X\sigma_Y\right)^2$
- We compared the results to a statistical simulation and obtained these results – a good match

	Analytic		Simulation	
	mean	sig	mean	sig
X	1.0000000	0.5000000	1.0000115	0.5000028
Y	1.0000000	0.5000000	1.0000026	0.4999844
XY	1.1000000	0.8789198	1.1038594	0.8839102

Product of Two Lognormal PDFs

- This is a more germane problem since products of RVs appear in factor CERs and they are typically the products of lognormal RVs
  - A lognormal distribution of the prime mission product (PMP) which is the assumed lognormal distribution of the sum of hardware and software costs, and
  - a cost factor with a lognormal error term
- This is a very difficult problem, believe it or not
  - We could find very little literature that provided the variance of the product of two bivariate lognormal distributions
  - We had to solve the problem explicitly
- The solution lies in dealing with the lognormal distributions as exponentiated normal distributions then using MOM on the normal parameters

The Square of Two Standard Lognormal PDFs, Slide 1 of 2

- The easiest example to test is the square of two lognormal variables, L(1,1)
  - This is a highly-skewed distribution
  - Since $\mu_X = 1$ and $\sigma_X = 1$ the math is very simple
  - The covariance is the product of the standard deviations,

$$\rho_{XX}\sigma_X\sigma_X = 1$$

  - The mean of the square of the distributions is

$$\mu_Z = \mu_X\mu_X + \rho_{XX}\sigma_X\sigma_X = (1)(1) + (1)(1)(1) = 2$$

The "flaw of averages"

The Square of Two Standard Lognormal PDFs, Slide 2 of 2

- Remembering that

$$Var[XY] = \rho_{X^2,Y^2}\sigma_{X^2}\sigma_{Y^2} + \cancelto{1}{\sigma_X{}^2}\mu_Y{}^2 + \cancelto{1}{\sigma_Y{}^2}\mu_X{}^2 + \cancelto{1}{\sigma_X{}^2}\sigma_Y{}^2 - [2\cancelto{2}{\rho_{X,Y}}\sigma_X\sigma_Y\mu_X\mu_Y + (\rho_{X,Y}\cancelto{1}{\sigma_X}\sigma_Y)^2]$$

  - $Var[X^2] = E[X^4] - E[X^2]E[X^2]$
- The expectations of X are
  - So $Var[X^2] = 64 - [2 * 2] = \mathbf{60}$
  - and $\sigma_{X^2} = \sqrt{Var[X^2]} = \sqrt{60} = 7.745966692$

$E(X^1)=$	1
$E(X^2)=$	2
$E(X^3)=$	8
$E(X^4)=$	64

- Comparing these results to a statistical simulation we get similar means but different sigmas (due to sampling of the tails)

	Analytic		Simulation	
	mean	sigma	mean	sigma
ZL	1.0000000	1.0000000	0.9998327	0.9963679
ZL2	2.0000000	7.7459667	1.9924044	7.0635770

The Variance of Products of RVs

- We can accept Goldberger's formula because it is not distribution-specific (i.e., it is not relevant to only bivariate normal distributions)
- This was a great discovery that allowed further work to be accomplished in the area of functional correlation, which we will discuss later

Agenda

- Introduction
- Mathematical Problems in Cost and Schedule Estimates
- Probability Tools
- Products of Random Variables
  - Expectation Methods
  - Propagation of Errors
- Functional Correlation
- Discrete Risks
- Max and Min of Random Variables
- Parametric Estimate Example Problem
- Resource-Loaded Schedule Example Problem
- Summary
- Future Research

Propagation of Errors

- The "Propagation of Errors" method allows us to calculate the mean and sigma values of the product of two *uncorrelated* random variables A and B

$$\mu_{AB} = \mu_A \mu_B$$

$$\sigma_{AB} = \sqrt{(\mu_A \sigma_B)^2 + (\sigma_A \mu_B)^2 + (\sigma_A \sigma_B)^2}$$

- Used to find moments of the product of two random independent variables such as a CER and its percent error
- The moments of the product AB do not rely on the shape of the distributions A and B, only their moments

Agenda

- Introduction
- Mathematical Problems in Cost and Schedule Estimates
- Probability Tools
- Products of Random Variables
    - Expectation Methods
    - Propagation of Errors
- Functional Correlation
- Discrete Risks
- Max and Min of Random Variables
- Parametric Estimate Example Problem
- Resource-Loaded Schedule Example Problem
- Summary
- Future Research

Functional Correlation, Slide 1 of 2

- What is it?
  - Functional correlation is the correlation induced by the functional relationship between RVs
- Who cares about it?
  - It is transparent in statistical simulations by their very nature
  - By defining a mathematical problem such as a schedule network or a cost estimate, the user only needs to correlate user-defined PDFs
  - In an analytic approach it is VERY important because the analyst must account for the relationships between RVs
- When does it occur?
  - When one RV is a function of another RV: $Y=f(X)$
  - When two RVs share a common RV: $Y_1=f_1(X)$, $Y_2=f_2(X)$
  - When two RVs share a dependent term: $Y_1=f_1(U)*e_1$, $Y_2=f_2(V)*e_2$ where $\rho_{e1,e2} \neq 0$ (i.e., their errors are correlated)

Functional Correlation, Slide 2 of 2

- These three types of correlations can be cast into three categories or types
  - Direct (nested) dependence of RVs
  - Shared dependence from use of the same RVs
  - Shared dependent RVs
- They can also be categorized in terms of how deeply they are related (their "order")

	Order 1	Order 2
Type I	$\rho_{X,Y}$ where $Y = f(X)$	$\rho_{X,Y}$ where $Y = f(g(X))$
Type II	ρ_{Y_1,Y_2} where $Y_1 = f_1(X)$ and $Y_2 = f_2(X)$	ρ_{Y_1,Y_2} where $Y_1 = f_1\big(g_1(X)\big)$ and $Y_2 = f_2(g_2(X))$
Type III	ρ_{Y_1,Y_2} where $Y_1 = f_1(X_1)\varepsilon_1$, $Y_2 = f_2(X_2)\varepsilon_2$, and $\rho_{\varepsilon_1,\varepsilon_2} \neq 0$ or $\rho_{X_1,X_2} \neq 0$	ρ_{Y_1,Y_2} where $Y_1 = f_1(g_1(X_1)\varepsilon_1)$, $Y_2 = f_2(g_2(X_2)\varepsilon_2)$, and $\rho_{\varepsilon_1,\varepsilon_2} \neq 0$, or $\rho_{X_1,X_2} \neq 0$

Steps to Compute Functional Correlation

1. Equate the correlation between two random variables

$$Corr(X,Y) = \rho_{X,Y} = \frac{E[XY] - \mu_X \mu_Y}{\sigma_X \sigma_Y}$$

2. Determine the components
   a) Find the means of the two RVs
   b) Find the variances of the two RVs
   c) Find the product of the two RVs
   d) Find the Expectation of 2c
3. Rewrite equation in terms of the components found in Steps 2a through 2d

This is why we call it Pearson “product moment” correlation

Functional Correlation Example 1

- Type II-1 Functional Correlation Example, $Y_i = (a_i + b_i x)\varepsilon_i$
  - Given two CERs, $Y_1 = (1 + 1.5x)\varepsilon_1$, and $Y_2 = (2 + 3x)\varepsilon_2$
  - If $X = T(2,3,4)$, $\sigma_{\varepsilon_1} = 0.2, \sigma_{\varepsilon_2} = 0.3$, and $\rho_{\varepsilon_1,\varepsilon_2} = 0.2$, then

$\mu_x = \frac{2+3+4}{3} = 3$, and $\sigma_X = 0.408248$ and $Var(X) = \sigma_X{}^2$=0.16667

$$\rho_{Y_1,Y_2} = \frac{\left(1 + \rho_{\varepsilon_1,\varepsilon_2}\sigma_{\varepsilon_1}\sigma_{\varepsilon_2}\right)(b_1 b_2\, E[x^{c_1+c_2}]) - \mu_{f_1}\,\mu_{f_2}}{\prod_{i=1}^{2}\left(\sqrt{\sigma_{f_i}{}^2 + \sigma_{\varepsilon_i}{}^2\mu_{f_i}{}^2 + \sigma_{f_i}{}^2\sigma_{\varepsilon_i}{}^2}\right)}$$

- Using values from above

$$\rho_{Y_1,Y_2} = \frac{(1.012)\,((1)(2)+3[(1)(3)+(2)(1.5)]+(1.5)(3)\,(9.16667))-60.5}{(1.2649)(3.5391)} = \frac{(1+0.012)(2+3[3+3]+41.25)-60.5}{(1.265)(3.539)}$$

$$\rho_{Y_1,Y_2} = \frac{(1.012)(2+18+41.25)-60.5}{4.4766} = \frac{61.985-60.5}{4.4766} = \frac{1.485}{4.4766} = 0.3317$$

Functional Correlation in Schedule Networks

- Functional correlation between two tasks T1 and T2 that have the same predecessor, P, that has a finish date F_P
  - Assume the durations of T1 and T2 (D_1 and D_2, respectively) are correlated by ρ_{D_1,D_2}
  - The start dates of T1 and T2 are F_1 and F_2 respectively
  - The finish dates of T1 and T2 are $F_1 = F_P + D_1$ and $F_2 = F_P + D_2$
- The correlation equation relies on the finish date of the predecessor, the covariance of the durations and the sigmas of F_1 and F_2

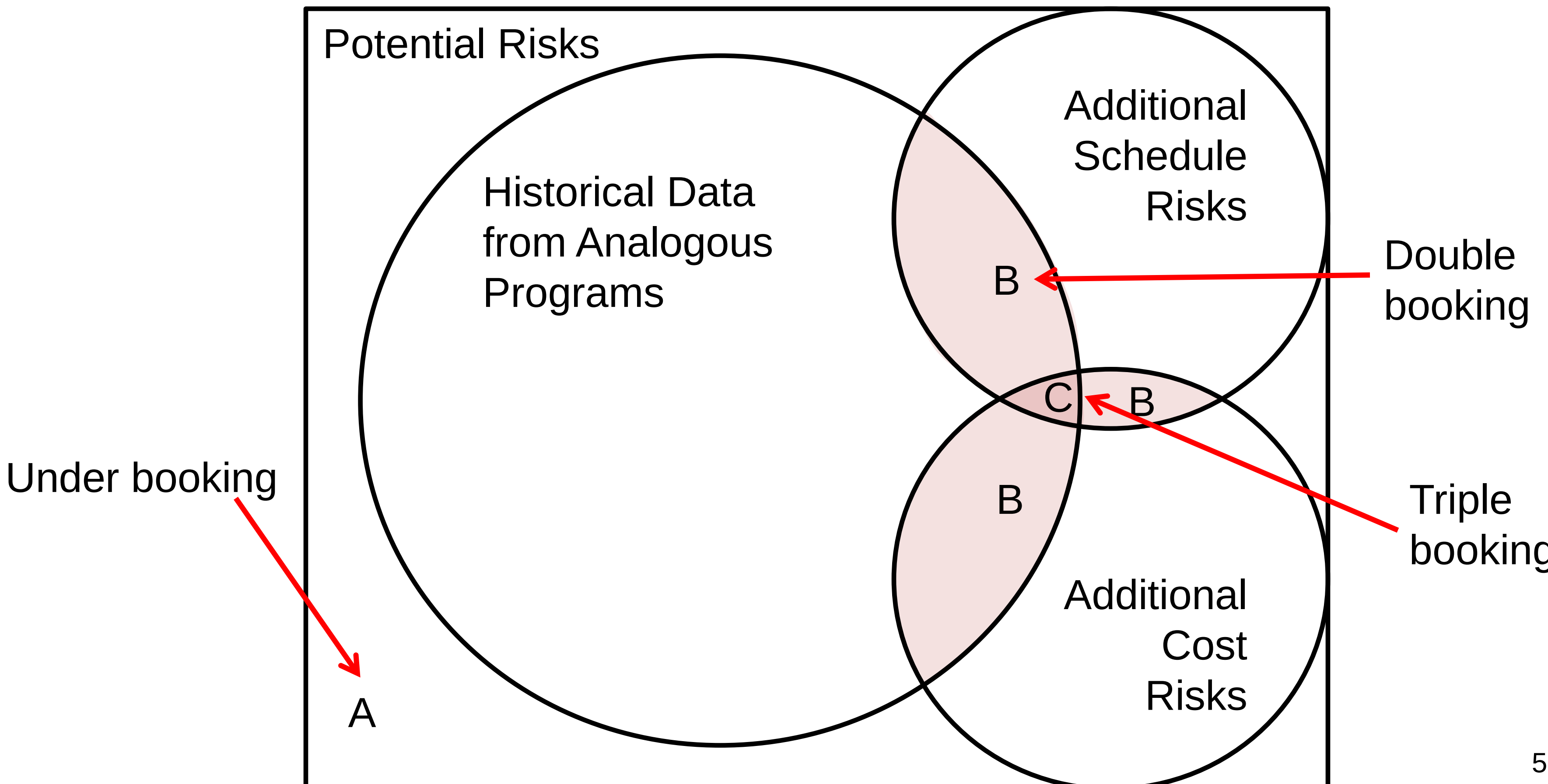
$$\rho_{F_1,F_2} = \frac{{\sigma_{F_P}}^2 + \rho_{D_1,D_2}\sigma_{D_1}\sigma_{D_2}}{\sqrt{{\sigma_{F_P}}^2 + {\sigma_{D_1}}^2}\sqrt{{\sigma_{F_P}}^2 + {\sigma_{D_2}}^2}}$$

Agenda

- Introduction
- Mathematical Problems in Cost and Schedule Estimates
- Probability Tools
- Products of Random Variables
  - Expectation Methods
  - Propagation of Errors
- Functional Correlation
- Discrete Risks
- Max and Min of Random Variables
- Parametric Estimate Example Problem
- Resource-Loaded Schedule Example Problem
- Summary
- Future Research

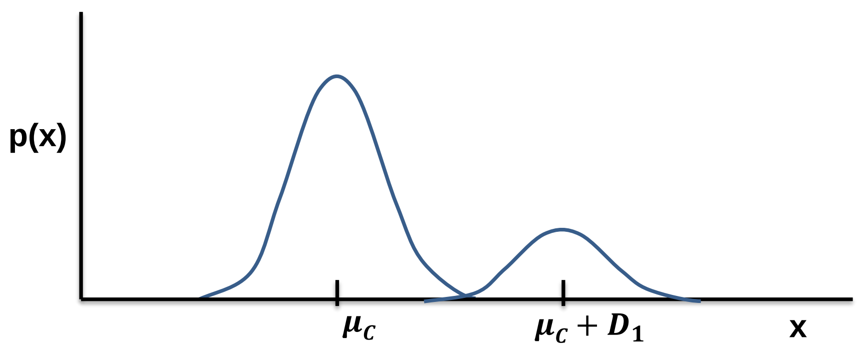
Discrete Risks

- Discrete risks should be included in a risk estimate
  - But it depends on the estimating methods used and their underlying data (i.e., more data results in estimates with more included risks)



Single Risk Case

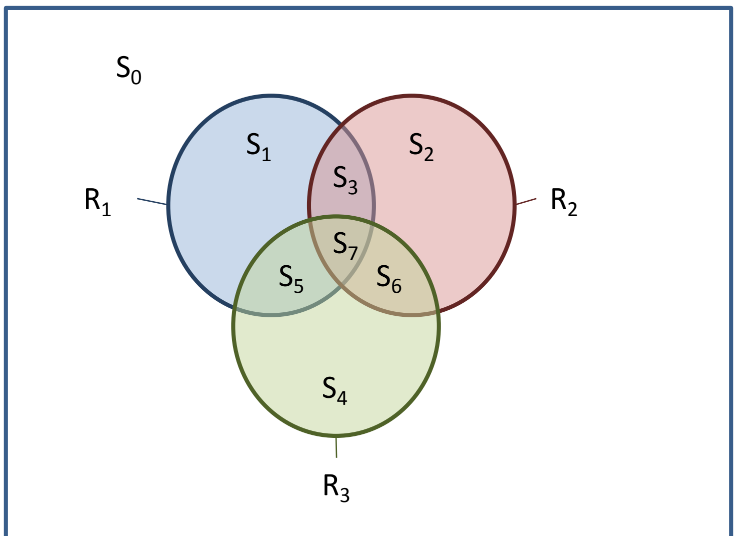
- When we combine a single discrete risk with an estimate represented by a continuous PDF, c, we get a mixed (or mixture) distribution
  - Discrete risk is defined by probability p_1 and impact D_1

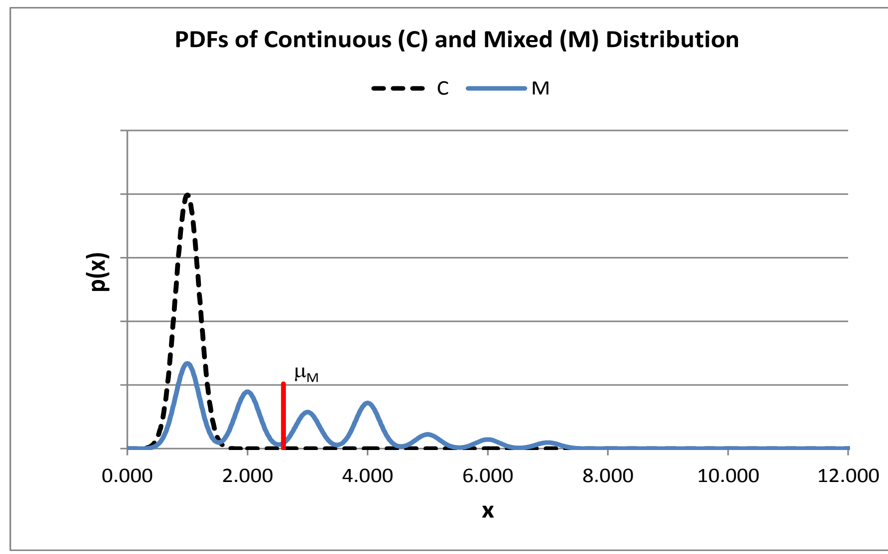


- The resulting mixed distribution is a probability-weighted distribution of the two risk states

Multiple Risks Case

- If we have n risks, then we will have 2^n possible risk states
- This forms a much more complicated distribution
  - The continuous distribution C is defined by a normal distribution, $N(1,0.2)$, and the three discrete risks are defined by $R_i(P_i, D_i)$: $R_1(0.4,1)$, $R_2(0.3,2)$, and $R_3(0.2,3)$





Moments of Mixed Distribution

- Fortunately, the report provides a method of calculating the moments of the mixed distribution

$$\mu_M = 1.0 + (0.4)(1.0) + (0.3)(2.0) + (0.2)(3.0) = 2.6$$

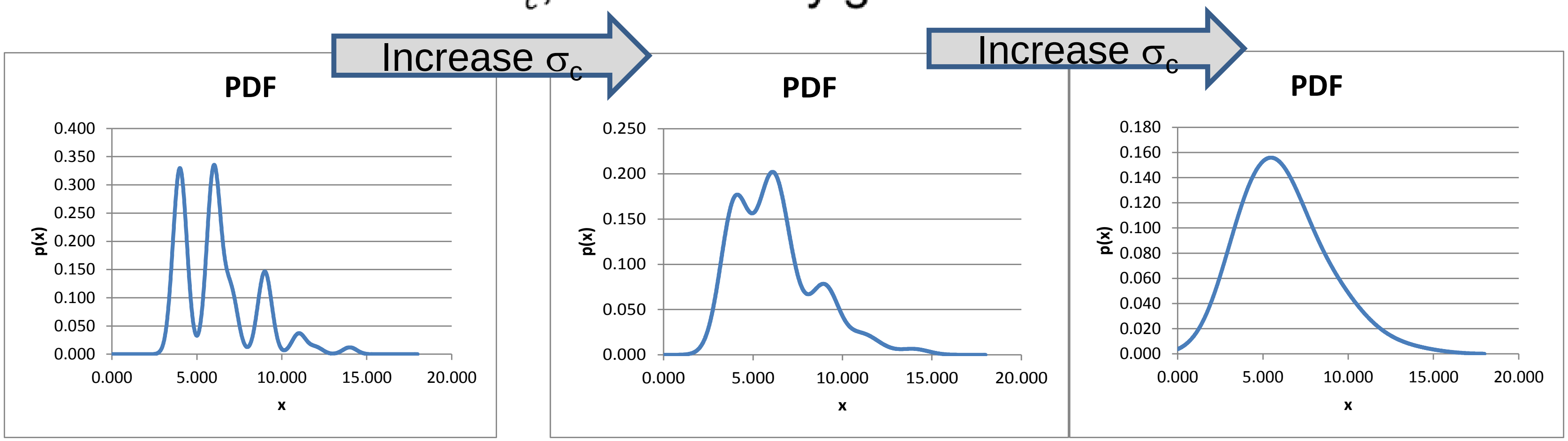
$$\sigma_M = \sqrt{\sum_{i=0}^{2^n-1} P(S_i)\left\{\left(\sigma_{D_{S_i}}\right)^2 + \left[D_{S_i} - \delta\mu\right]^2\right\}} = 1.6460$$

- Comparison with Simulated Results
  - Minor differences due to simulation's inability to exactly draw correlated random variables

Exact	Simulated
$\mu_M = 2.6000$	$\hat{\mu}_M = 2.6004$
$\sigma_M = 1.6460$	$\hat{\sigma}_M = 1.6495$

Discrete Risk Example

- Assume a cost estimate, C, has a normal distribution with $\mu_c = 4.0$ and $\sigma_c = 0.4082$ and the following three risks (R(prob., impact)) are identified: $R_1(0.50, 2)$, $R_2(0.25, 3)$, $R_3(0.10, 5)$
- Then the mixed PDF looks like this:
- And as we increase σ_c, we instantly get:



- We can change risks R, R, R or add/remove them easily

Many Discrete Risks

- The case where there are many discrete risks is VERY interesting
- The mixed distribution appears to be a much better representation of fat-tailed distributions
  - And better yet, we can calculate the moments of the mixed distribution
  - And we can determine the contributions of the estimate's underlying distribution and the discrete risks forming it

Agenda

- Introduction
- Mathematical Problems in Cost and Schedule Estimates
- Probability Tools
- Products of Random Variables
  - Expectation Methods
  - Propagation of Errors
- Functional Correlation
- Discrete Risks
- Max and Min of Random Variables
- Parametric Estimate Example Problem
- Resource-Loaded Schedule Example Problem
- Summary
- Future Research

Minimum and Maximum of Random Variables

- Absolutely necessary when performing analytic schedule risk assessment (SRA) due to mathematics of merge points of parallel tasks
- We know the exact PDF and moments of the max of two correlated Gaussian distributions but VERY little work can be cited that solves the problem of the max of correlated non-Gaussian distributions
  - The author's recent SSCAG paper was a big step forward, but it does not have the capability of dealing with highly correlated RVs
  - We are still working on more solutions
- For the time being, we know how to compute the max and min of two Gaussian distributions and limited cases of all others

Max and Min of Lognormal PDFs

- The first two moments of the maximum of two lognormal distributions (used to compare max finish dates in a schedule network)
  - Correlation and overlap of distributions cause the mean to shift

$$E[X] =$$

$$\mu_1 \Phi\left[\frac{(P_1 - P_2) + (Q_1{}^2 - \rho Q_1 Q_2)}{\theta}\right] + \mu_2 \Phi\left[\frac{(P_2 - P_1) + (Q_2{}^2 - \rho Q_1 Q_2)}{\theta}\right]$$

$$E[X^2] = (\sigma_1^2 + \mu_1^2)\Phi\left(\frac{P_1 - P_2}{\theta}\right) + (\sigma_2^2 + \mu_2^2)\Phi\left(\frac{P_2 - P_1}{\theta}\right)$$

$\theta = \sqrt{Q_1^2 + Q_2^2 - 2\rho Q_1 Q_2}$ where the correlation between their underlying normal distributions is

$\rho = \frac{1}{Q_1 Q_2} ln\left[1 + \rho_{1,2}\left(\sqrt{[e^{Q_1^2} - 1][e^{Q_2^2} - 1]}\right)\right]$, and

$\rho_{1,2}$ = Pearson correlation between lognormal distributions of tasks X_1 and X_2

P_1, P_2, Q_1,and Q_2 are parameters of the lognormal distribution defined in Equations 4-5 and 4-6.

Agenda

- Introduction
- Mathematical Problems in Cost and Schedule Estimates
- Probability Tools
- Products of Random Variables
  - Expectation Methods
  - Propagation of Errors
- Functional Correlation
- Discrete Risks
- Max and Min of Random Variables
- Parametric Estimate Example Problem
- Resource-Loaded Schedule Example Problem
- Summary
- Future Research

Example Cost Model

- Consider this simple parametric cost model ($\rho_{\varepsilon i,\varepsilon j} = 0.2$)

	WBS Element, i	CER, i	Drivers	X_i	ε_i
1	Systems Engineering, Program Management Integration and Test (SEITPM)	$Y_1 = 0.498 X_1{}^{0.9}\varepsilon_1$	PMP	$\sim L\left(\sum_{i=2}^{10}\mu_i, \sqrt{\sigma^T\rho\sigma}\right)$	L(1,0.49)
	Prime Mission Product (PMP)	$\sum_{i=2}^{10} Y_i$	Sum of Hardware and Software costs		0
2	Antenna	$Y_2 = 34.36 X_{2a}{}^{0.5} X_{2b}{}^{0.8}\varepsilon_2$	Aperture Diameter (m), Frequency (GHz)	T(2,3,4) T(16,17,18)	L(1,0.30)
3	Electronics	$Y_3 = 30.06 X_3{}^{0.8}\varepsilon_3$	Frequency (GHz)	T(16,17,18)	L(1,0.40)
4	Platform	$Y_4 = 26.91 X_{4a}{}^{0.5} X_{4b}{}^{0.85}\varepsilon_4$	Aperture Diameter (m), Number of Axes	T(2,3,4) Constant = 2	L(1,0.38)
5	Facilities	$Y_5 = 1.64 X_5{}^{0.8}\varepsilon_5$	Area (m^2)	T(18,20,22)	L(1,0.25)
6	Power Distribution	$Y_6 = 0.32 X_6{}^{0.9}\varepsilon_6$	Electrical Power (W)	T(1200,1425,1875)	L(1,0.18)
7	Computers	$Y_7 = 0.58 X_7{}^{0.87}\varepsilon_7$	MFLOPS	T(180,200,220)	L(1,0.31)
8	Environmental Control	$Y_8 = 1.94 X_8{}^{0.4}\varepsilon_8$	Heat Load (W)	T(1100,1200,1300)	L(1,0.21)
9	Communications	$Y_9 = 5.62 X_9{}^{0.9}\varepsilon_9$	Data Rate (MBPS)	T(25,30,35)	L(1,0.28)
10	Software	$Y_{10} = 1.38 X_{10}{}^{1.2}\varepsilon_{10}$	Effective Source Lines of Code, eKSLOC	T(80,90,130)	L(1,0.32)

FRISK Method?

- We will apply the FRISK method to solve this, but we must first:
  - Find the distributions of the WBS elements
  - Find the correlation matrix between the WBS elements
- Finding the distributions of WBS elements 2-10 using expectation methods is the easy part

WBS #	Analytic		Simulation	
	μ	σ	μ	σ
2	572.706	177.022	572.676	176.900
3	289.953	116.136	289.962	116.172
4	83.829	32.484	83.824	32.463
5	18.014	4.544	18.014	4.543
6	230.920	46.015	230.911	45.977
7	58.248	18.186	58.244	18.172
8	33.068	6.960	33.068	6.959
9	119.965	34.446	119.962	34.420
10	347.121	120.764	347.121	120.787

Cost-on-Cost Functions

- Finding the distribution of WBS element 1 (the cost-on-cost function) is not so simple
  - First we have to determine the functional correlations between WBS elements 2-10, find the moments of that sum (the prime mission product, or “PMP”), then calculate the moments of WBS element 1 using expectation methods
- The functional correlation matrix (FCM) has these types of correlations

Cost-on-cost

Shared variables

Correlated errors

$\rho_{yi,yj}$	1	2	3	4	5	6	7	8	9	10
1	1.0000	I-2	I-2	I-2	I-2	I-2	I-2	I-2	I-2	I-2
2	I-2	1.0000	II-1	II-1	III-1	III-1	III-1	III-1	III-1	III-1
3	I-2	II-1	1.0000	III-1	III-1	III-1	III-1	III-1	III-1	III-1
4	I-2	II-1	III-1	1.0000	III-1	III-1	III-1	III-1	III-1	III-1
5	I-2	III-1	III-1	III-1	1.0000	III-1	III-1	III-1	III-1	III-1
6	I-2	III-1	III-1	III-1	III-1	1.0000	III-1	III-1	III-1	III-1
7	I-2	III-1	III-1	III-1	III-1	III-1	1.0000	III-1	III-1	III-1
8	I-2	III-1	III-1	III-1	III-1	III-1	III-1	1.0000	III-1	III-1
9	I-2	III-1	III-1	III-1	III-1	III-1	III-1	III-1	1.0000	III-1
10	I-2	III-1	III-1	III-1	III-1	III-1	III-1	III-1	III-1	1.0000

Computing Moments of WBS Element 1

- After computing the FCM sub-matrix of the PMP elements, we can use the FRISK method to find the mean and standard deviation of PMP
- We then use that to find the moments of WBS element 1

WBS #	Analytic		Simulation	
	μ	σ	μ	σ
1	413.170	201.048	413.090	200.916
2	572.706	177.022	572.676	176.900
3	289.953	116.136	289.962	116.172
4	83.829	32.484	83.824	32.463
5	18.014	4.544	18.014	4.543
6	230.920	46.015	230.911	45.977
7	58.248	18.186	58.244	18.172
8	33.068	6.960	33.068	6.959
9	119.965	34.446	119.962	34.420
10	347.121	120.764	347.121	120.787

- Then we can statistically sum all of the WBS elements

Applying the FRISK Method

- Next we have to find how WBS element 1 is functionally correlated to the rest of the WBS elements (shown below)

$\rho_{yi,yj}$	1	2	3	4	5	6	7	8	9	10
1	1.0000	0.2614	0.2098	0.1454	0.1156	0.1426	0.1273	0.1184	0.1393	0.2085
2	0.2614	1.0000	0.1969	0.2306	0.1924	0.1753	0.1927	0.1937	0.1893	0.1785
3	0.2098	0.1969	1.0000	0.1959	0.1979	0.1804	0.1983	0.1993	0.1948	0.1837
4	0.1454	0.2306	0.1959	1.0000	0.1944	0.1772	0.1947	0.1957	0.1912	0.1804
5	0.1156	0.1924	0.1979	0.1944	1.0000	0.1790	0.1968	0.1978	0.1933	0.1823
6	0.1426	0.1753	0.1804	0.1772	0.1790	1.0000	0.1794	0.1803	0.1762	0.1662
7	0.1273	0.1927	0.1983	0.1947	0.1968	0.1794	1.0000	0.1981	0.1936	0.1827
8	0.1184	0.1937	0.1993	0.1957	0.1978	0.1803	0.1981	1.0000	0.1946	0.1836
9	0.1393	0.1893	0.1948	0.1912	0.1933	0.1762	0.1936	0.1946	1.0000	0.1794
10	0.2085	0.1785	0.1837	0.1804	0.1823	0.1662	0.1827	0.1836	0.1794	1.0000

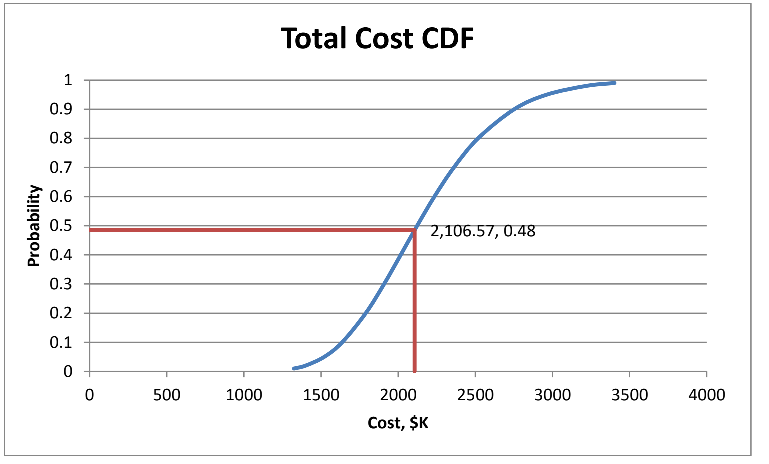
- Now that we have the moments of each WBS element and a full FCM, we can calculate the statistics of the total using FRISK (whew!)

	Analytic		Simulation	
	µ	σ	µ	σ
Total	2166.995	443.915	2166.873	443.511

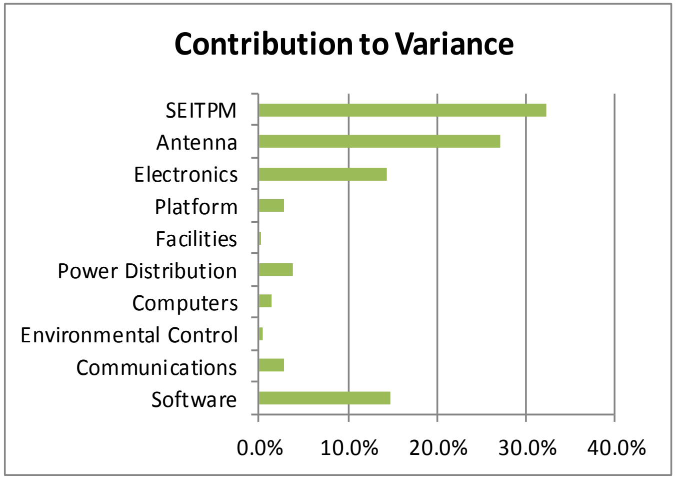
The Results

- We can now plot the CDF, show percentiles, and each WBS element's contribution to the variance of the total

Percentile	Total Cost, Y
10%	1637.140582
20%	1789.878287
30%	1908.780462
40%	2016.616222
50%	2122.909227
60%	2234.804788
70%	2361.059157
80%	2517.905056
90%	2752.814045

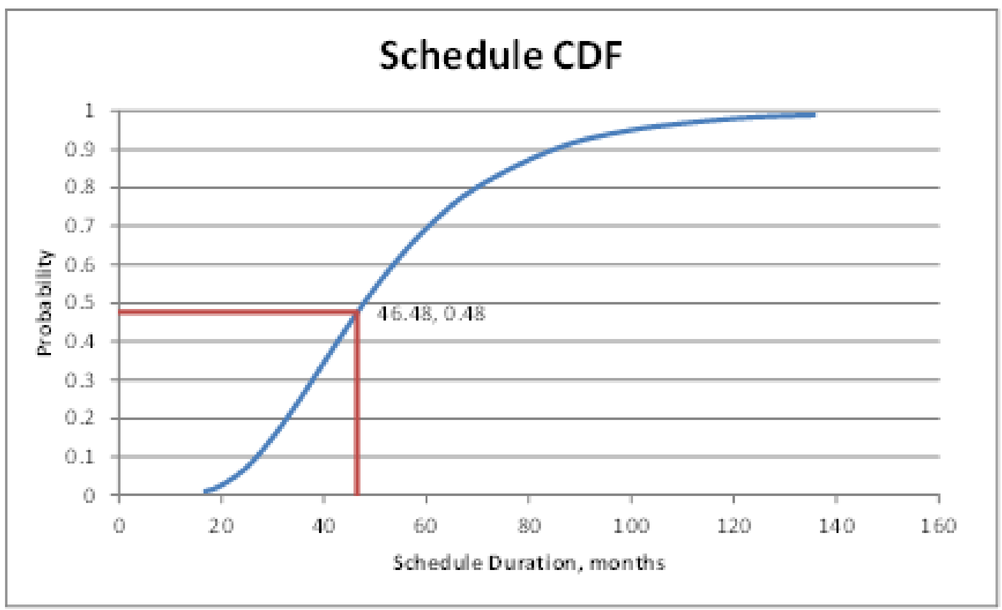


WBS	CTV
1 SEITPM	32%
2 Antenna	27%
3 Electronics	14%
4 Platform	3%
5 Facilities	0%
6 Power Distribution	4%
7 Computers	1%
8 Environmental Control	0%
9 Communications	3%
10 Software	15%
Sum	1.0000



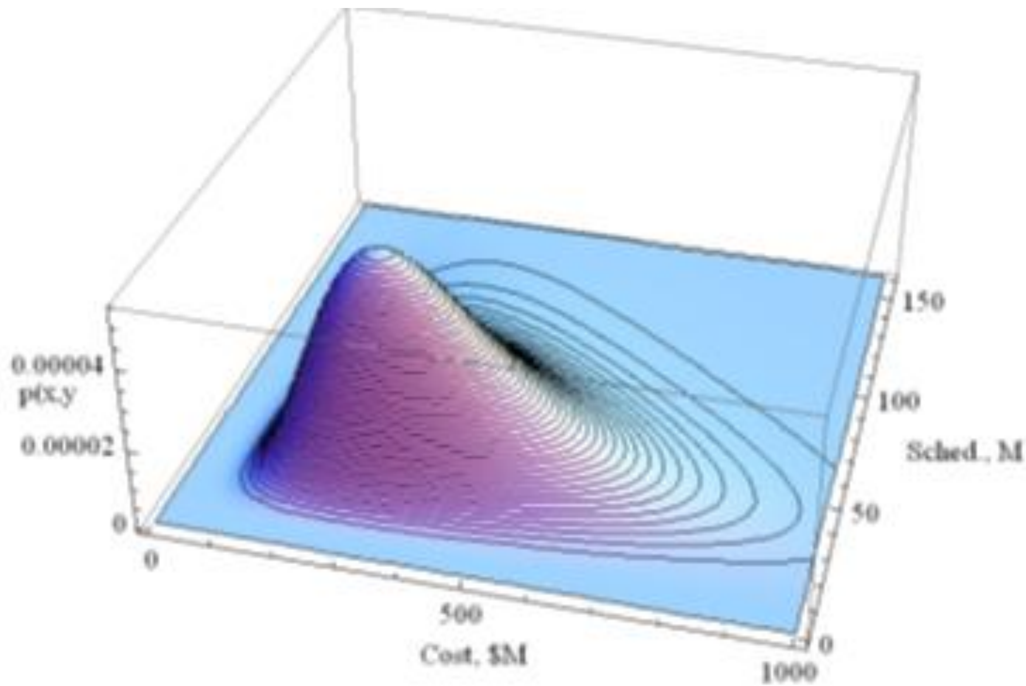
Schedule Distribution

- We also modeled the schedule as a PDF based on a fictitious schedule estimating relationship (SER): $D = 0.21{X_D}^{1.2}\varepsilon_D$.
- The schedule will now be correlated to the cost of the program through the dependent variable, X_D, which is "effective source lines of code"
  - The moments are $\mu_D = 52.823$, and $\sigma_D = 24.935$.



Joint Cost and Schedule Distribution

- The joint PDF of cost and schedule is modeled as a bivariate lognormal distribution; where cost and schedule are linked by a Type II-2 functional correlation, $\rho_{Y,D} = 0.0364$
  - The value $\rho_{Y,D}$ calculated from a 100,000-trial statistical simulation is 0.0366, which indicates excellent agreement with the analytic result



Agenda

- Introduction
- Mathematical Problems in Cost and Schedule Estimates
- Probability Tools
- Products of Random Variables
  - Expectation Methods
  - Propagation of Errors
- Functional Correlation
- Discrete Risks
- Max and Min of Random Variables
- Parametric Estimate Example Problem
- Resource-Loaded Schedule Example Problem
- Summary
- Future Research

Resource-Loaded Schedule

- NASA provided a resource-loaded schedule with which to perform a demonstration of the analytic method
- This consisted of a MS Project file
- The MS Project file was exported to MS Excel for analysis
- Two problems needed to be solved in order to perform an analytical assessment on the resource-loaded schedule:
    - How to deal with calendar days (contiguous) versus working days (non-contiguous) in the analysis
    - How to formulate the schedule into a linearized sum of durations
    - How to calculate the correlation between cost and schedule

Calendar Days and Work Days

- When a duration of a task causes it finish to occur on the following day, schedule programs can quickly determine which WORKDAY the task will finish
- Simple addition of durations that result in finish dates (that are otherwise consecutive calendar days) will provide incorrect dates (non-working days)
- In a resource-loaded schedule, we are interested in schedule finish dates that are calculated in the correct calendar days, but we need the number of workdays to allow for resource loading
- We performed the schedule analysis using addition and subtraction of days (to simplify the resource-loading math) then converted the schedule results to account for working days using the "WORKDAY" function in Excel

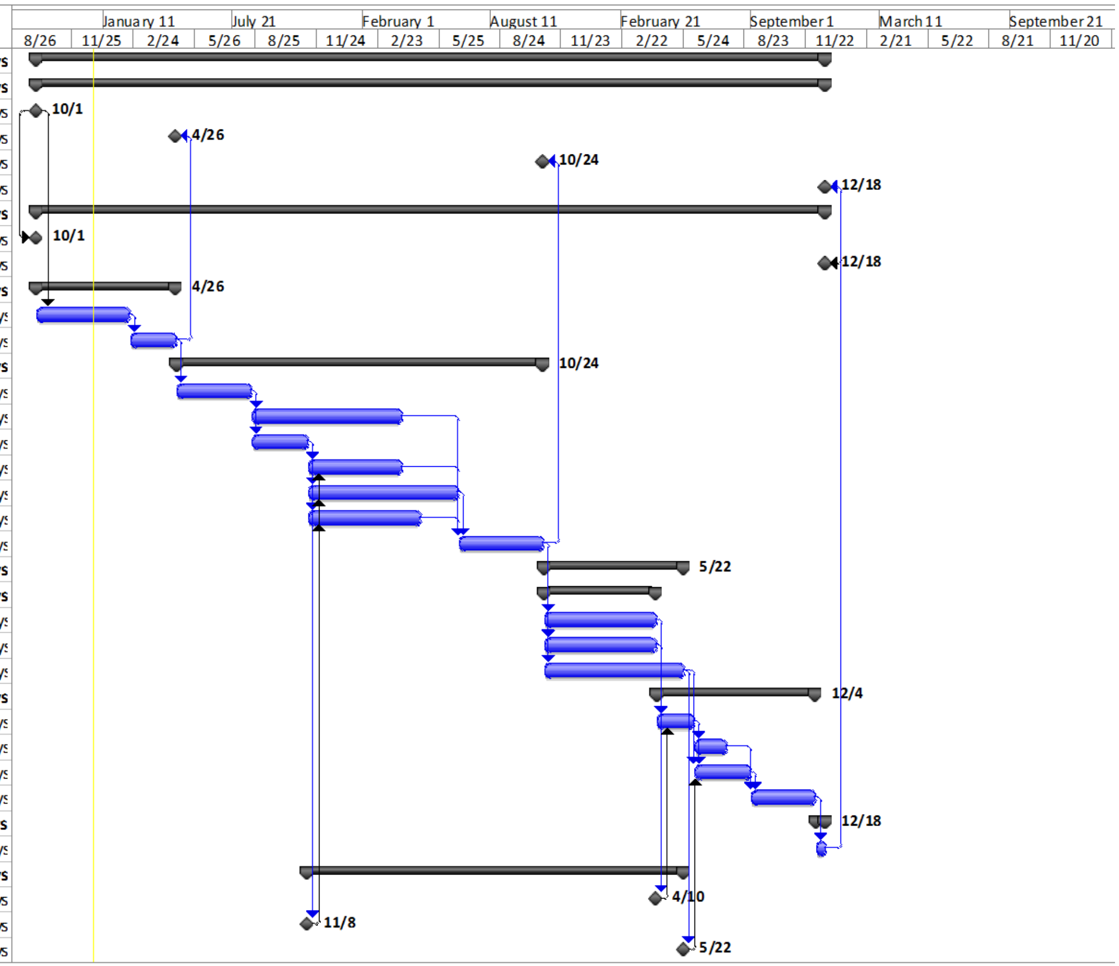
Calculating Probabilistic Task Durations

- Probabilistic task durations were calculated based on the parameters specified in the MS Project file
- The distributions consisted of additive and multiplicative distributions that were either triangular, normal, lognormal or discrete
- Three of the duration PDFs were correlated
- Two of the durations were defined as discrete risks
- Using the methods outlined in the report, we could quickly and easily determine the moments of these durations
- We discuss the issue of how to model the schedule network

Schedule Network Model, Slide 1 of 3

- The schedule network looks like this:

ID	Task Name	Duration
1	**JACS Analysis File**	**840 days**
2	**Milestone Summary**	**840 days**
3	Project ATP	0 days
4	PDR	0 days
5	CDR	0 days
6	Rocket delivery	0 days
7	**Project Support Costs hammock task**	**840 days**
8	Support Start	0 days
9	Support Finish	0 days
10	**Preliminary Design**	**150 days**
11	Requirements definition and documentation	100 days
12	Preliminary design activities	50 days
13	**Detailed Design**	**390 days**
14	Initial detailed design	80 days
15	Design GN&C	160 days
16	Trade studies and analysis	60 days
17	Design pyrotechnics	100 days
18	Design propulsion system	160 days
19	Design structures and mechanisms	120 days
20	Finalize integrated design	90 days
21	**Development and Unit Testing**	**150 days**
22	**Fabricate rocket Components**	**120 days**
23	Fabricate and unit test structure (includin	120 days
24	Fabricate and unit test engine	120 days
25	Develop and test flight software for GN&C	150 days
26	**Integration and Testing**	**170 days**
27	Integrate rocket components	40 days
28	Test frame, fuel system and engine	35 days
29	Test guidance system	60 days
30	Final integration and testing	70 days
31	**Delivery**	**10 days**
32	Delivery	10 days
33	**Risk Register**	**400 days**
34	Risk 1 - TI - Additional Purchase	0 days
35	Risk 2 - Duration - Additional Studies Required	0 days
36	Risk 3 - TI and Duration - Delay from Additional Sof	0 days



SEER by GALORATH

Covarus Business · Math · Science

Schedule Network Model, Slide 2 of 3

- We reduced this to a "linearized schedule" representing the duration of the project
  - $D = D_{11} + D_{12} + D_{14} + D_{[15,19]} + D_{20} + D_{[23,29]} + D_{30} + D_{32}$, where
  - $D_{[15,19]} = \max(D_{15}, D_{16} + D_{35} + D_{17}, D_{16} + D_{35} + D_{18}, D_{16} + D_{35} + D_{19})$, and
  - $D_{[23,29]} = max\{max[max(D_{23}, D_{24}) + D_{27}, D_{25} + D_{36}] + D_{29}, max(D_{23}, D_{24}) + D_{27}$
  - The finish (F_i) dates of these tasks were computed by adding the ATP date to the sum of durations up to that of the particular task of interest
- Since tasks 11, 12 and 14 were serial and task 14's finish date was lognormal, these finish dates were all lognormally distributed – which was a blessing and a curse
  - We know how to compute max(A,B) where A and B are normal, but

Schedule Network Model, Slide 3 of 3

- In two instances (i.e., at merge points) we had to calculate the max finish date of tasks:
  - One where durations were correlated
  - One where durations were assumed independent
- In both cases, we had to determine the functional correlation of the finish dates of the merging tasks because they shared the same predecessor
  - The variance of the finish date of the predecessor linked their finish date variances
- When compared to a 100,000 trial statistical simulation the results were an excellent match

	Analytic		Simulation	
	mean	sigma	mean	sigma
max(15,17,18,19)	01/10/14	16.02	01/10/14	16.0638
max(23,24)	08/08/14	19.69	08/08/14	19.75134
max(27,36)	09/28/14	20.78	09/27/14	21.55239
max(28,29)	11/30/14	21.67	11/28/14	22.43632

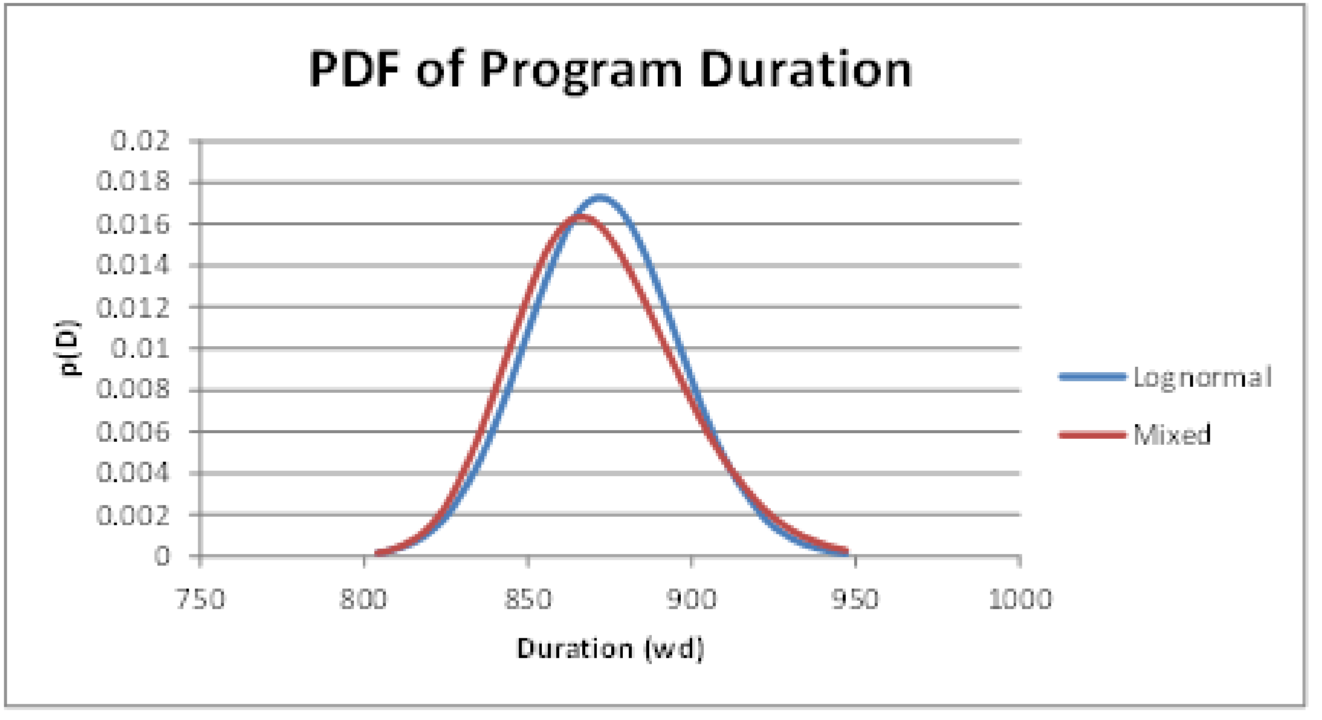
Schedule Risk Results

- Once we knew the moments of the finish dates of the tasks, calculating the finish date statistics of the project was fairly simple
- Comparing the results of our analytic approximation to a 100,000-trial statistical simulation we see very good agreement as well.
  - Differences in the statistics are due to sampling errors in the simulation (for wd statistics) and due to conversion of the analytic results into calendar dates (for cd statistics).

Finish Date	Analytic Approach	Statistical Simulation
$\mu_{F'}$ (wd)	02/20/15	02/18/15
$\sigma_{F'}$ (wd)	23.09	23.74
μ_F (cd)	02/05/16	01/24/16
σ_F (cd)	32.34	33.17

Schedule Duration

- The schedule duration was modeled using a single lognormal distribution and a mixed distribution representing the inclusion of discrete risks



Criticality Index (1)

- The criticality index (CI) was computed based on the probability a particular task is on the critical path
  - In other words, The CI is the probability that one task's finish date is greater than others, or Prob(X>Y)
  - This can be re-written as Prob(Y<X)=Prob(Y-X<0)
- So we want to find the area of the PDF of Y-X from -∞ to 0
  - Its mean will be $E[Y-X]=E[Y]-E[X]$
  - Its variance will be $Var[Y-X]=Var[X]+Var[Y]-2Cov(X,Y)$
- Since all of our merging tasks had lognormally distributed finish dates, this was a fairly simple task
- It allowed us to calculate the CI of each schedule element and linearize the schedule a little more

Criticality Index (2)

- CI of tasks in NASA schedule network

Task ID	$\mu_{Duration}$	$\sigma_{Duration}$	μ *Start*	σ *Start*	μ *Finish*	σ_{Finish}	CI
10	213.50	4.88	10/01/12	0.00	05/02/13	4.88	100%
11	142.33	4.37	10/01/12	0	02/20/13	4.37	100%
12	71.17	2.18	02/20/13	4.37	05/02/13	4.88	100%
14	115.73	6.98	05/02/13	4.88	08/26/13	8.52	100%
15	231.47	13.97	08/26/13	8.52	04/14/14	16.36	0%
16	86.80	5.24	08/26/13	8.52	11/21/13	10.00	100%
17	144.67	8.73	11/26/13	17.57	04/20/14	19.62	0%
18	231.47	13.97	11/26/13	17.57	07/16/14	22.45	100%
19	173.60	10.48	11/26/13	17.57	05/19/14	20.46	0%
20	130.20	7.86	07/16/14	22.44	11/23/14	23.78	100%
23	159.60	14.55	11/23/14	23.78	05/01/15	27.88	16.62%
24	159.60	14.55	11/23/14	23.78	05/01/15	27.88	16.62%
25	220.50	10.50	11/23/14	23.78	07/01/15	25.99	66.76%
27	56.00	8.40	05/06/15	27.58	07/01/15	28.83	33.24%
28	50.63	5.03	07/01/15	28.83	08/20/15	29.26	0%
29	86.80	8.63	07/16/15	29.10	10/10/15	30.35	100%
30	100.80	10.34	10/13/15	30.35	01/22/16	32.06	100%
32	14.00	4.20	01/22/16	32.06	02/05/16	32.34	100%
34	0.00	0.00	05/01/15	27.88	05/01/15	27.88	0%
35	5.60	14.45	11/21/13	10.00	11/26/13	17.57	100%
36	10.50	16.12	07/01/15	25.99	07/12/15	30.58	66.76%

Linearized Schedule

- The results of the CI analysis showed which tasks needed to be considered in the linearized schedule
- Those tasks with CI=0 were eliminated
- Our linearized schedule becomes

$$D = D_{11} + D_{12} + D_{14} + D_{16} + D_{35} + D_{18} + D_{20} + D_{[23,28]} + D_{29} + D_{30} + D_{32}, \text{ where } D_{[23,28]} = max\left[max(D_{23}, D_{24}) + D_{27}, D_{25} + D_{36}\right]$$

Cost of Resource-Loaded Schedule

- The cost of WBS elements were linked to task durations in the schedule network
- Individual lowest-level WBS element Costs, X_i, are defined by the combination of TD* and TI* costs as follows:

$$X_i = \begin{cases} (TD_i \varepsilon_{TD_i})(TI_i \varepsilon_{TI_i}) = Duration'_i \varepsilon_{TD_i} Rate_i \varepsilon_{TI_i} & , if\ TI\ is\ multiplicative \\ [(TD_i \varepsilon_{TD_i})(TI_i)] + \varepsilon_i = (Duration'_i \varepsilon_{TD_i} Rate_i) + \varepsilon_{TI_i} & , if\ TI\ is\ additive \end{cases},$$

where:

ε_{TI_i} is the TI PDF

ε_{TD_i} is the TD PDF

$Duration'_i$ is the probabilistic task duration in wd.

$Rate_i$ is the nominal cost per wd.

- The individual task costs were statistically summed to their respective summary-level WBS elements using Method of Moments (FRISK)

* Time-dependent (TD) and time-independent (TI)

Discrete Risks

$Y = aX_1^{\,b} X_2^{\,c}\,\varepsilon$, where

Y is cost, a random variable (RV)

a, b, and c are constants, $a = 0.1$, $b = 0.95$, and $c = 0.60$

X_1 is a cost driver that is a RV, $X_1 = T(9,10,15)$

X_2 is a cost driver that is a RV, $X_2 = T(30,40,60)$

ε is the percent standard error of the CER, a RV, $\varepsilon = N(1,0.3)$

WBS	MC mean	MC sigma	MOM Cost mean	MOM Cost sigma
2	**$7,899,158.10**	**$8,793,491.21**	**$7,900,000.00**	**$8,783,507.27**
2.1	$3,097,397.91	$4,764,566.91	$3,100,000.00	$4,768,647.61
2.2	$0.00	$0.00		
2.3	$4,801,760.19	$7,379,295.72	$4,800,000.00	$7,376,313.44

Probabilistic Weighting

- To create the mixed distribution of the project cost, $f_{X_m}(x)$, we combine the continuous and discrete distributions
  - $f_{X_m}(x) = \sum_{i=0}^{3} p_{S_i} f_{X_{S_i}}(x)$, where
  - $f_{X_m}(x)$ represents the probability-of-occurrence-weighted sum of the individual states' PDFs
  - p_{S_i} = the probability of occurrence of state S_i
  - $f_{X_{S_i}}(x)$ = the PDF of state S_i

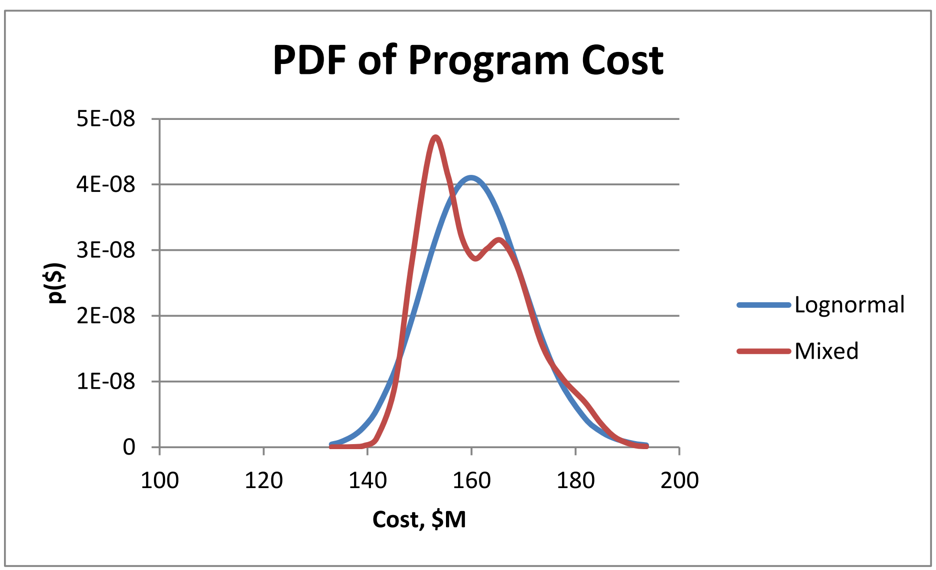
State	Prob.	μ_X	σ_X
S_0	0.49	\$152,860,068.75	\$4,272,695.15
S_1	0.21	\$163,193,402.08	\$4,394,482.83
S_2	0.21	\$168,860,068.75	\$4,519,136.03
S_3	0.09	\$179,193,402.08	\$4,634,452.08

Results

- The analytic results of the total cost were compared to those from a 100,000 trial statistical simulation, and the results agreed rather well

	Computed Values		Difference	
	Analytic	Staitistical Simulation	Additive	Percent
Mean	$160,810,256.90	$160,756,334.56	($53,922.34)	-0.034%
Sigma	$9,765,611.10	$10,053,584.87	($287,973.77)	-2.949%

- The PDFs of the lognormal approximation and the exact, mixed distribution are

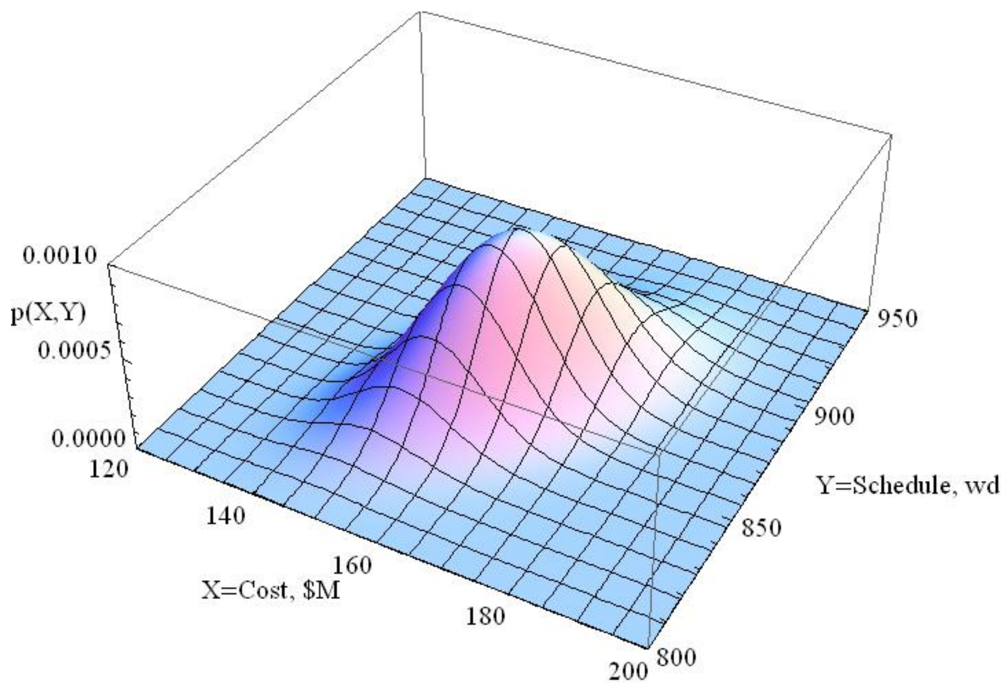


Joint Cost-Schedule PDF

- The joint cost-schedule PDF was developed using the marginal distributions of cost and schedule using the analytic results provided earlier
- The joint PDF was modeled as a bivariate lognormal distribution
  - This required knowing how the marginal distributions of cost and schedule were correlated
  - Fortunately, the correlation terms were vastly simplified by the structure of the schedule
  - $\rho_{X,D} = \frac{E[XD]-E[X]E[D]}{\sigma_X \sigma_{D'}} = \frac{E[XD]-\mu_X \mu_{D'}}{\sigma_X \sigma_{D'}}$ where $X = \sum_{i=LLWBS} X_i$, the sum of the costs of the lowest-level WBS elements, X_i
  - $XD = \left(\sum_{i=EL} X_i\right)\left(D_{11} + D_{12} + D_{14} + D_{[15,19]} + D_{20} + D_{[23,29]} + D_{30} + D_{32}\right)$

The Joint PDF

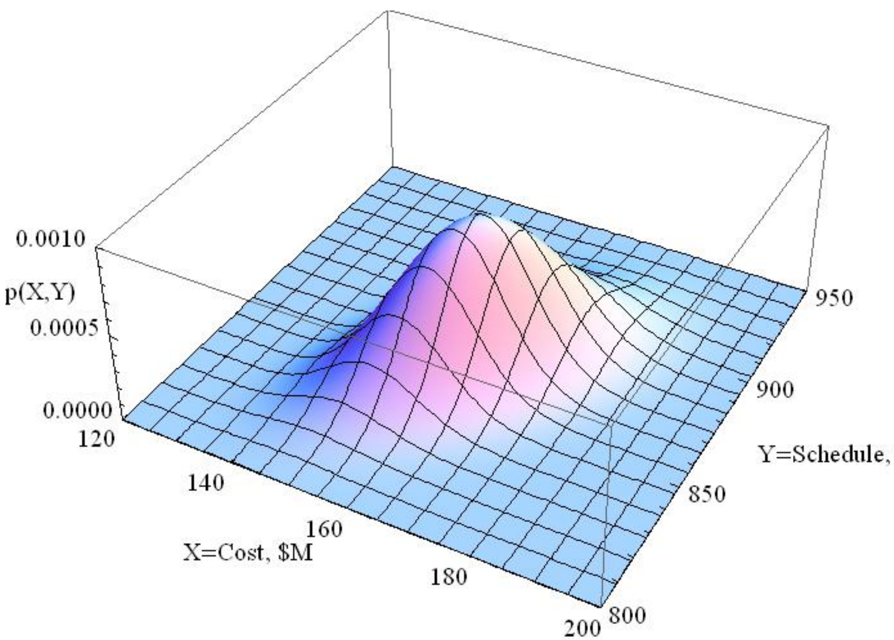
- The resulting calculations show $\rho_{X,D} = 0.5322$
  - The results from the 100,000-trial statistical simulation show $\rho_{X,D} = 0.5597$, which is very similar
- The resulting joint PDF is pictured below



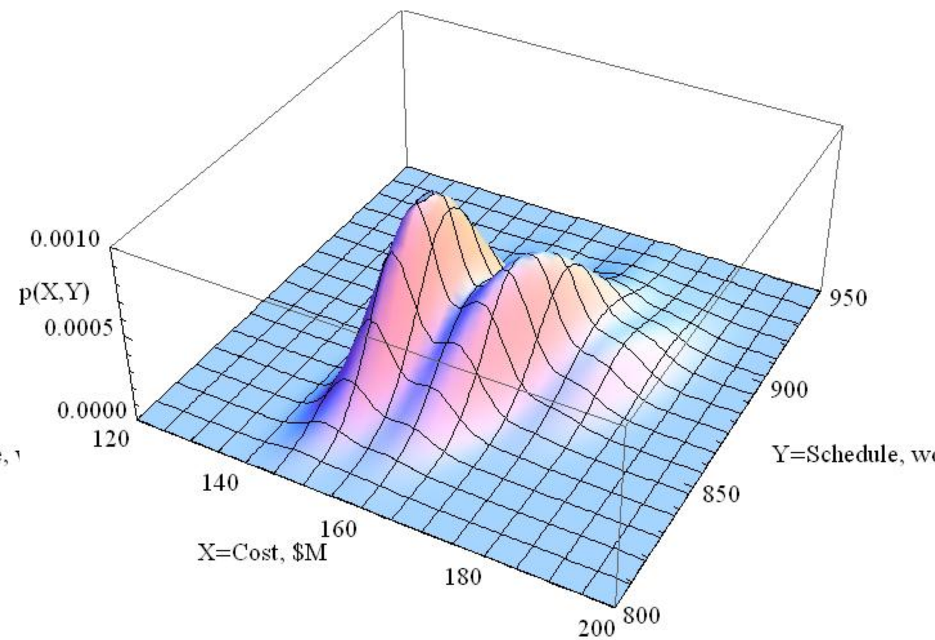
Comparing Distribution Plots

- The mixed distribution provides more graphical information

Bivariate Lognormal



Mixed Bivariate Lognormal



- The joint probability of meeting the point estimates of cost ($151M) and schedule (840 wd) is about 1.3%

Agenda

- Introduction
- Mathematical Problems in Cost and Schedule Estimates
- Probability Tools
- Products of Random Variables
    - Expectation Methods
    - Propagation of Errors
- Functional Correlation
- Discrete Risks
- Max and Min of Random Variables
- Parametric Estimate Example Problem
- Resource-Loaded Schedule Example Problem
- Summary
- Future Research

Summary

- Analytic techniques like MOM can be used in a variety of cost and schedule estimating problems
- By their nature, they require more work up-front to create, but they provide the exact moments and do it instantly
  - The payoff is when we reuse the models
- There have been gaps in the literature surrounding several aspects of MOM for which we have provided solutions, but we are confident there will be others!
- In the near future, expect to see analytic techniques applied more and more to cost and schedule risk analysis problems

Agenda

- Introduction
- Mathematical Problems in Cost and Schedule Estimates
- Probability Tools
- Products of Random Variables
    - Expectation Methods
    - Propagation of Errors
- Functional Correlation
- Discrete Risks
- Max and Min of Random Variables
- Parametric Estimate Example Problem
- Resource-Loaded Schedule Example Problem
- Summary
- Future Research

Future Research

- Evaluating Statistical Simulations
  - How do we know our simulations are accurate? Test them.
- Using Estimating Methods
  - Do CERs provide better estimates over other methods when discrete risks are considered? I think so.
- Basis of Estimate Credibility
  - Multiple actuals provide a better case for an estimate, and should instill more confidence in a bid. How do we prove this?
- Developing Cost Models
  - Can we perform errors-in-variables (EIV) regression as easily as regression of discrete variables? I think so!
- Improving Cost and Schedule Risk Tools
  - How can MOM be effectively used in our cost models? MOM, MOM/sim. Hybrids?
- Time-Phasing a Resource Loaded Schedule
  - How do you include time-phasing in a resource-loaded schedule? Very possible to do.

References (1)

1. Archibald, R. D., & Villoria, R. L. (1967). Network-Based Management Systems (PERT/CPM). New York: John Wiley & Sons.
2. Balanda, K., & MacGillivray, H. (1988). Kurtosis: A critical review. American Statistician, 42, 111-119.
3. Bohrnstedt, G. W., & Goldberger, A. S. (1969, Dec.). On the Exact Covariance of Products of Random Variables. Journal of the American Statistical Association, 64(328), 1439-1442.
4. Book, S. A. (1994). Do Not Sum 'Most Likely' Cost Estimates. 1994 NASA Cost Estimating Symposium. Houston, TX.
5. Book, S. A. (2011). Schedule Risk Analysis: Why It is Important and How to Do It. 2011 ISPA/SCEA Joint Annual Conference & Training Workshop. Albuquerque, NM: ISPA/SCEA.
6. Coleman, R. J., & Gupta, S. S. (1994). An Overview of Correlation and Functional Dependencies in Cost Risk and Uncertainty Analysis. 1994 DoD Cost Analysis Symposium. Williamsburg, VA.
7. Covert, R. P. (2001). Correlation Coefficients in the Unmanned Space Vehicle Cost Model Version 7 (USCM 7) Database. 3rd Joint ISPA/SCEA International Conference. Tyson's Corner, VA.
8. Covert, R. P. (2002). Comparison of Spacecraft Cost Model Correlation Coefficients. SCEA National Conference. Scottsdale, AZ.
9. Covert, R. P. (2006). “Errors-In-Variables Regression”. Joint SSCAG/EACE/SCAF Meeting. London, UK.
10. Covert, R. P. (2006). A Guide to Probabilistic Cost Estimation for Trade Studies. Crystal Ball User’s Conference. Denver, CO.
11. Covert, R. P. (2006). A Simplified Approach to Cost Evaluation in Source Selections. 2006 ISPA Conference. Seattle, WA.
12. Covert, R. P. (2006). Correlations in Cost Risk Analysis. 2006 Annual SCEA Conference. Tysons Corner, VA.

References (2)

13. Covert, R. P. (2008). Reducing Systemic Errors in Cost Models. 2008 ISPA/SCEA Joint International Conference. Noordwijk, NE.
14. Covert, R. P. (2011). Using Method of Moments in Schedule Risk Analysis. Bethesda, MD: IPM.
15. Covert, R. P., & Anderson, T. P. (2002). Regression of Cost Dependent CERs. 2002 SCEA National Conference. Phoenix, AX.
16. Darlington, R. B. (1970). Is Kurtosis Really "Peakedness?". American Statistician, 24, 19-22.
17. DeCarlo, L. (1997). On the meaning and use of kurtosis. Psychological Methods, 292-307.
18. Evans, M. J., & Rosenthal, J. S. (2010). Probability and Statistics: The Science of Uncertainty, 2nd Ed. New York, NY: W. H. Freeman and Co.
19. Flynn, B., Braxton, P., Garvey, P., & Lee, R. (2012). Enhanced Scenario-Based Method for Cost Risk Analysis: Theory, Application and Implementation. 2012 SCEA/ISPA Joint Annual Conference & Training Workshop. Orlando, FL.
20. Garvey, P. R. (1999). Do Not Use Rank Correlation in Cost Risk Analysis. 32nd DOD Cost Analysis Symposium.
21. Garvey, P. R. (2000). Probability Methods for Cost Uncertainty Analysis: A Systems Engineering Perspective. New York, NY: Marcel Dekker.
22. Goodman, L. A. (1960, Dec.). On the Exact Variance of Products. Journal of the American Statistical Association, 55(292), 708-713.
23. Iman, R., & Conover, W. (1982). A Distribution-free Approach to Inducing Rank Correlation among Input Variables. Communications in Statistics - Simulation, Computation, Vol. 11, 311-334.
24. Kal, S. (2002). Basic Electronics: Devices, Circuits and IT Fundamentals. New Delhi, India: Prentice Hall.

References (3)

25. Lien, D. (2005). On the Minimum and Maximum of Bivariate Lognormal Random Variables. Extremes, 8, 79-83.
26. Lo, C. F. (2012, May 22). The Sum and Difference of Two Lognormal Random Variables. Hong Kong, China. Retrieved December 12, 2012, from SSRN: http://ssrn.com/abstract=2064829 or http://dx.doi.org/10.2139/ssrn.2064829
27. Lurie, P., & Goldberg, M. (1998, Feb). A Method for Simulating Correlated Random Variables from Partially Specified Distributions. Management Science, 44(2), 203-218.
28. MacCrimmon, K. R., & Ryavec, C. A. (1962). An Analytical Study of the Pert Assumptions. Santa Monica, CA: RAND.
29. MacKenzie, D., & Addison, B. (2000). Space System Cost Variance and Estimating Uncertainty. 70th SSCAG Meeting, Boeing Training Center. Tukwila, WA.
30. Moors, J. (1986). The meaning of kurtosis: Darlington reexamined. American Statistician, 40, 283-284.
31. Nadarajah, S., & Kotz, S. (2008, Feb.). Exact Distribution of the Max/Min of Two Gaussian Random Variables. IEEE Transactions on VLSI Systems, 16(2), 210-212.
32. Papoulis, A. (1965). Probability, Random Variables and Stochastic Processes. New York, NY: McGraw Hill.
33. Young, P. H. (1992). FRISK- Formal Risk Assessment of System Cost. Aerospace Design Conference. Irvine, CA: AIAA.